Icahn Institute: Stay connected with us!

@multiscalebio

Mining the digital universe of data to develop personalized cancer therapies

August 12, 2013

Icahn School of Medicine at
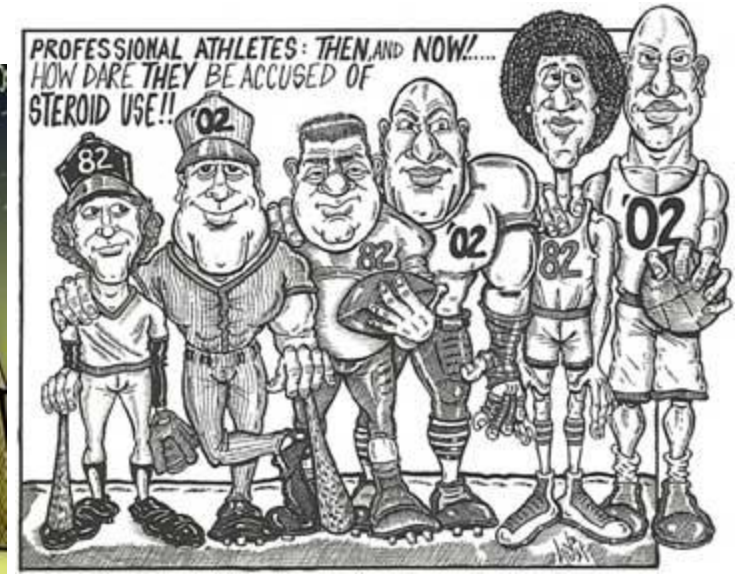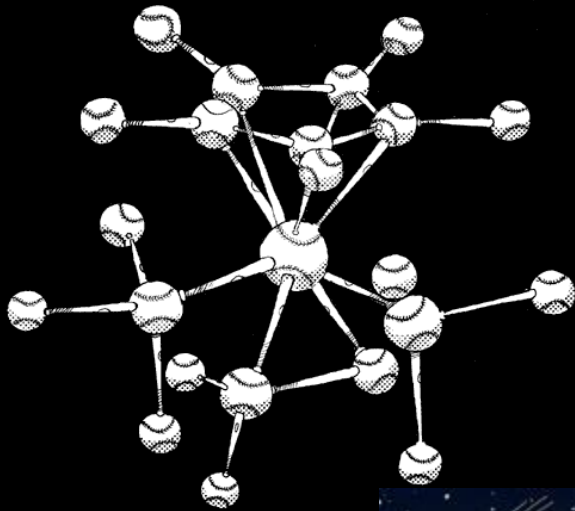**Mount Sinai**

Disclosures

- I am on the Scientific Advisory Board for
  - Pacific Biosciences
  - Numedii
  - StationX
  - Spiral Genetics
  - Berg Pharmaceuticals
  - Ingenuity
  - GNS Healthcare

- I am on the Board of Directors for
 - Sage Bionetworks
  - While Biome

Disclosures

- Given the apparent rampant use of performance enhancing drugs in sports:

Poll results: look who's doping

In January, *Nature* launched an informal survey into readers' use of cognition-enhancing drugs. **Brendan Maher** has waded through the results and found large-scale use and a mix of attitudes towards the drugs.
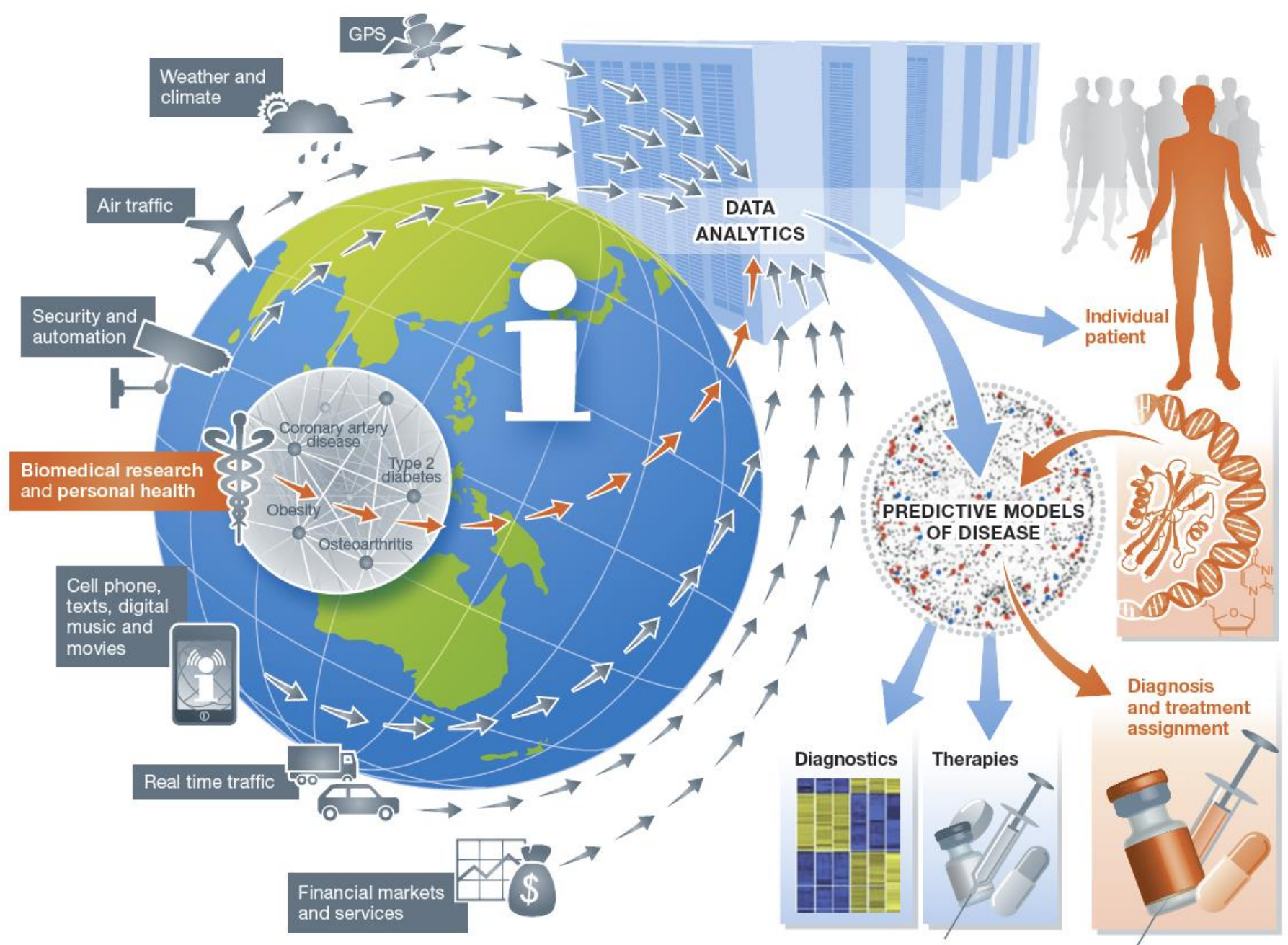

TRENDS IN USE OF NEUROENHANCERS
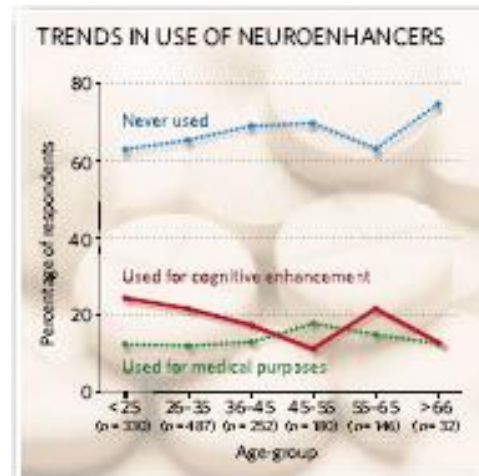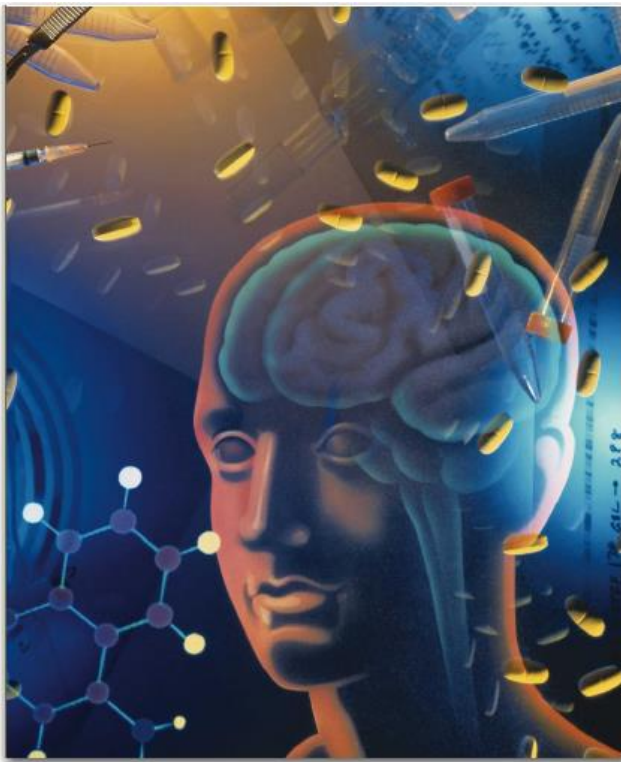
I fall on the "never used" curve

20% are abusing

- I used no performance enhancing drugs to carry out any of the research I will discuss today

Considering the digital universe of data to better diagnose and treat patients

We need to be able to leverage the digital universe of information to best solve the most challenging problems

1.8
ZETTABYTES

(1.8 trillion gigabytes of information will be created and replicated in 2011; growth continues to accelerate – factor of 9 growth in last 5 years)

Being masters of really big data now critical for biomedical research (TB → PB → EB → ZB)
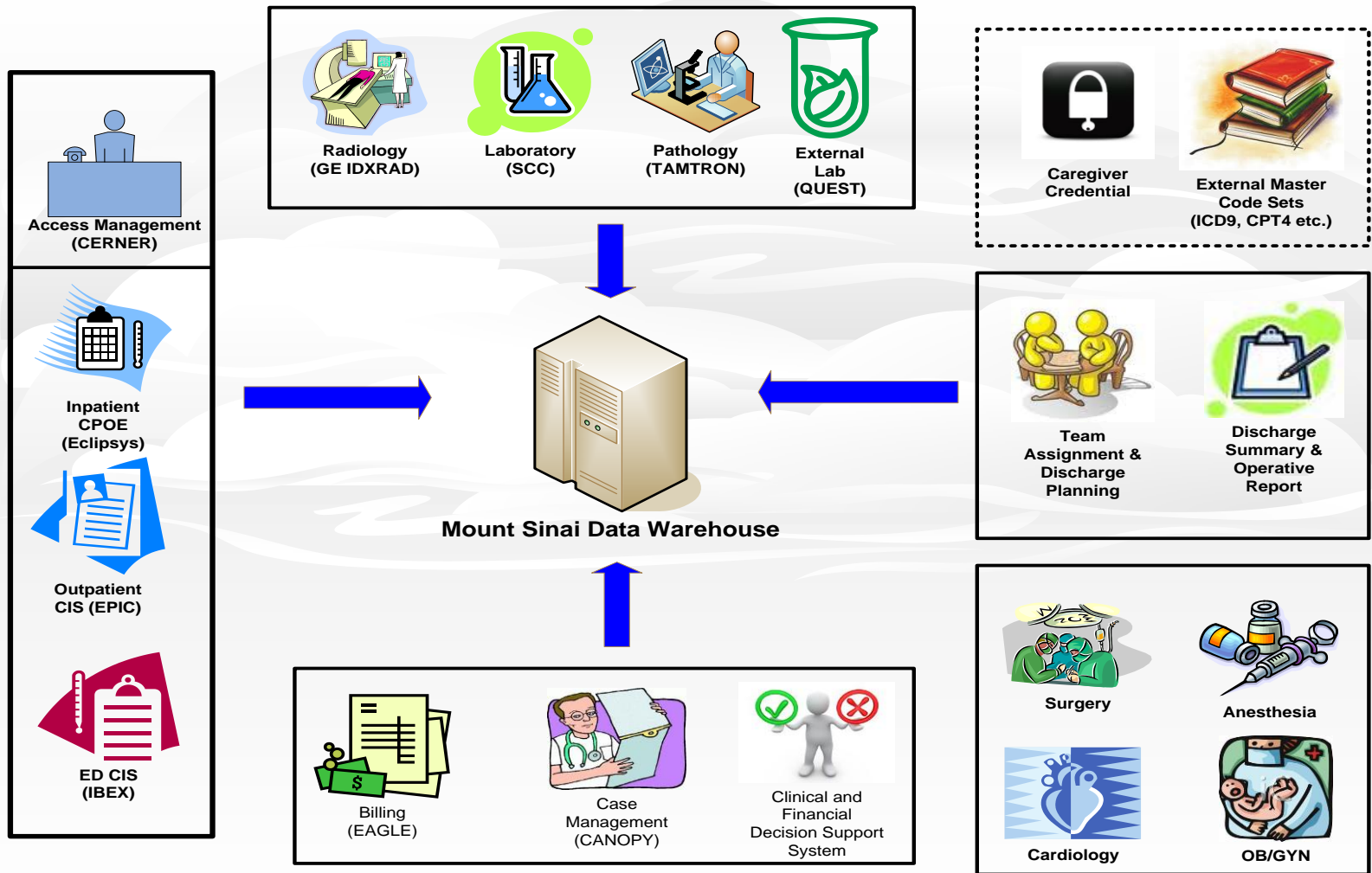

Cost per Raw Megabase of DNA Sequence
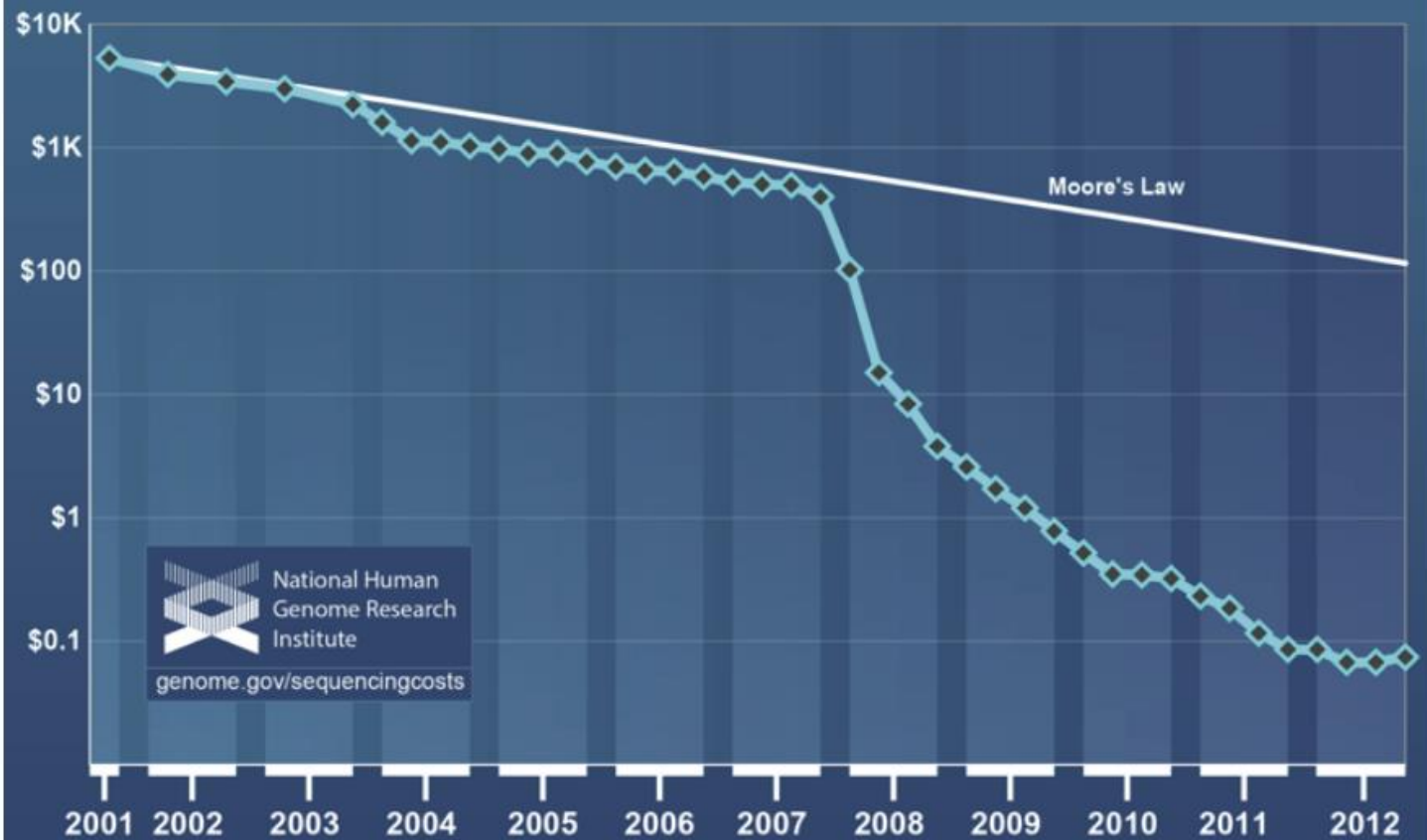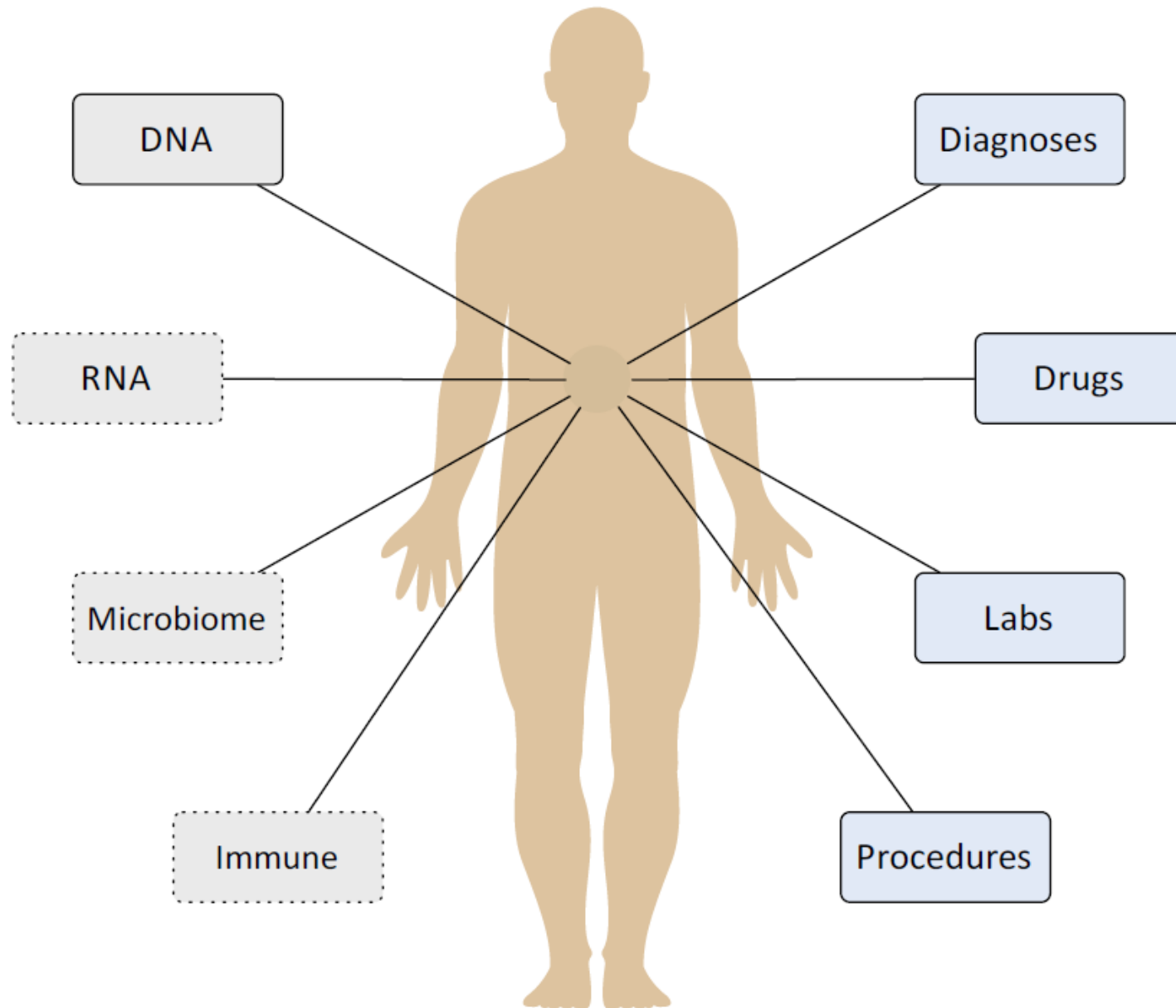
Organisms — Tissues — Single cells — Single cell, real-time, continuous?

Big Data Warehouses at Medical Centers like Mount Sinai Contain Virtually All Facts And Transaction Records For Millions of Patients
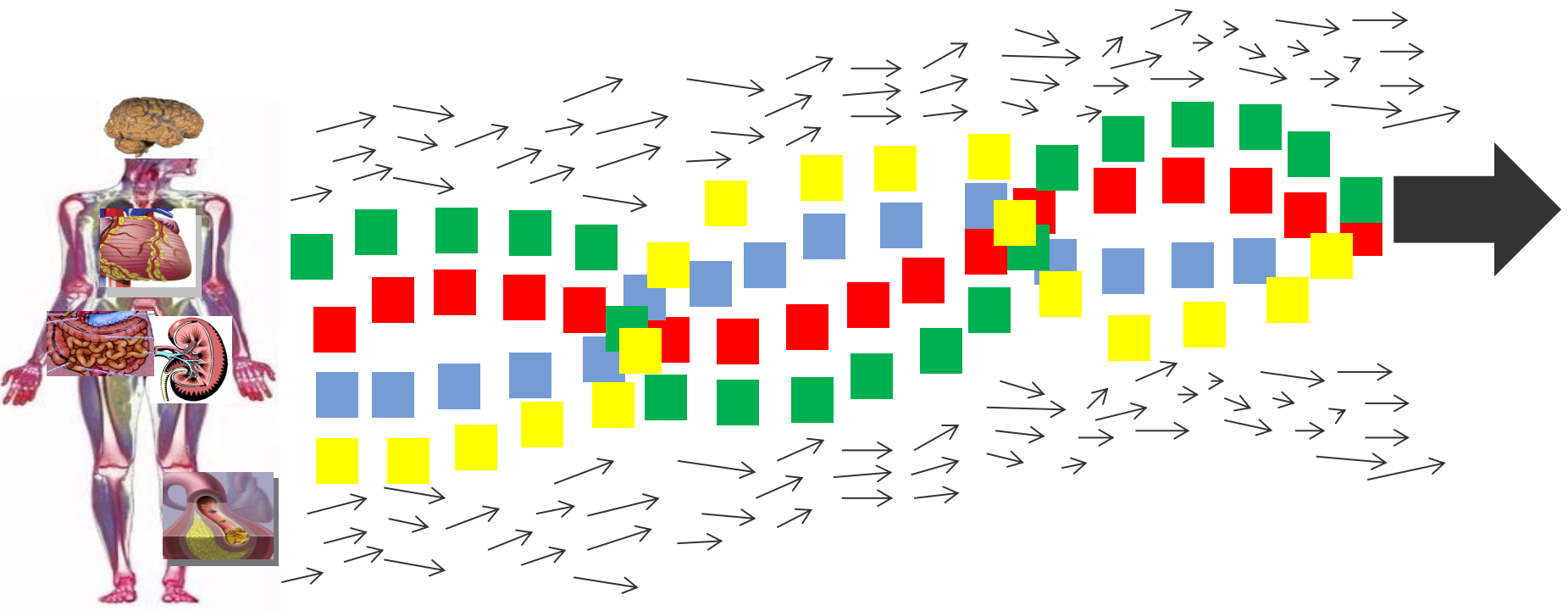

Institute for Personalized Medicine at Mount Sinai

Multiscale measures of patients now available through efforts like Mount Sinai's Biobank (>25,000 *identified* patients and growing fast)

Modified and unmodified DNA
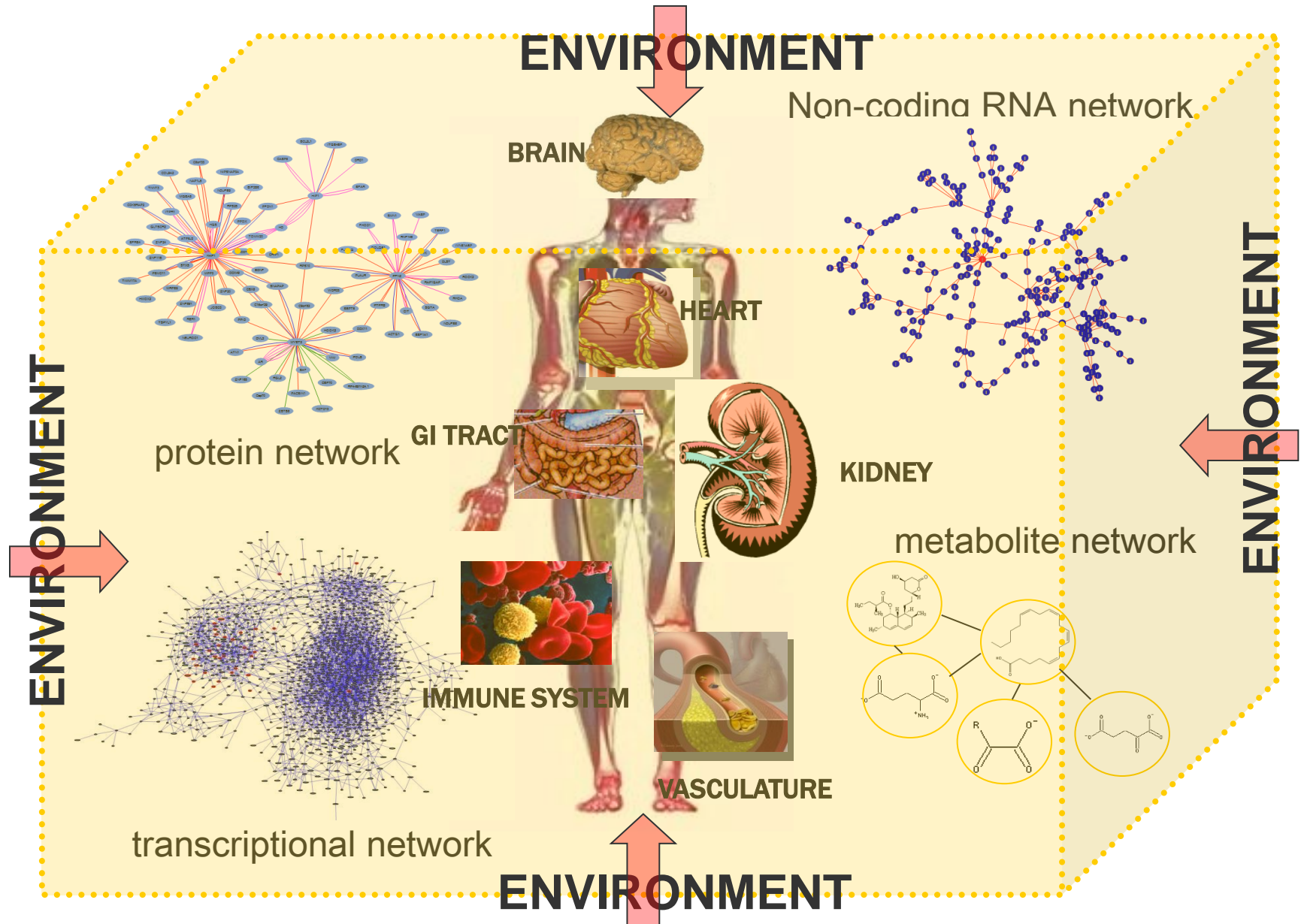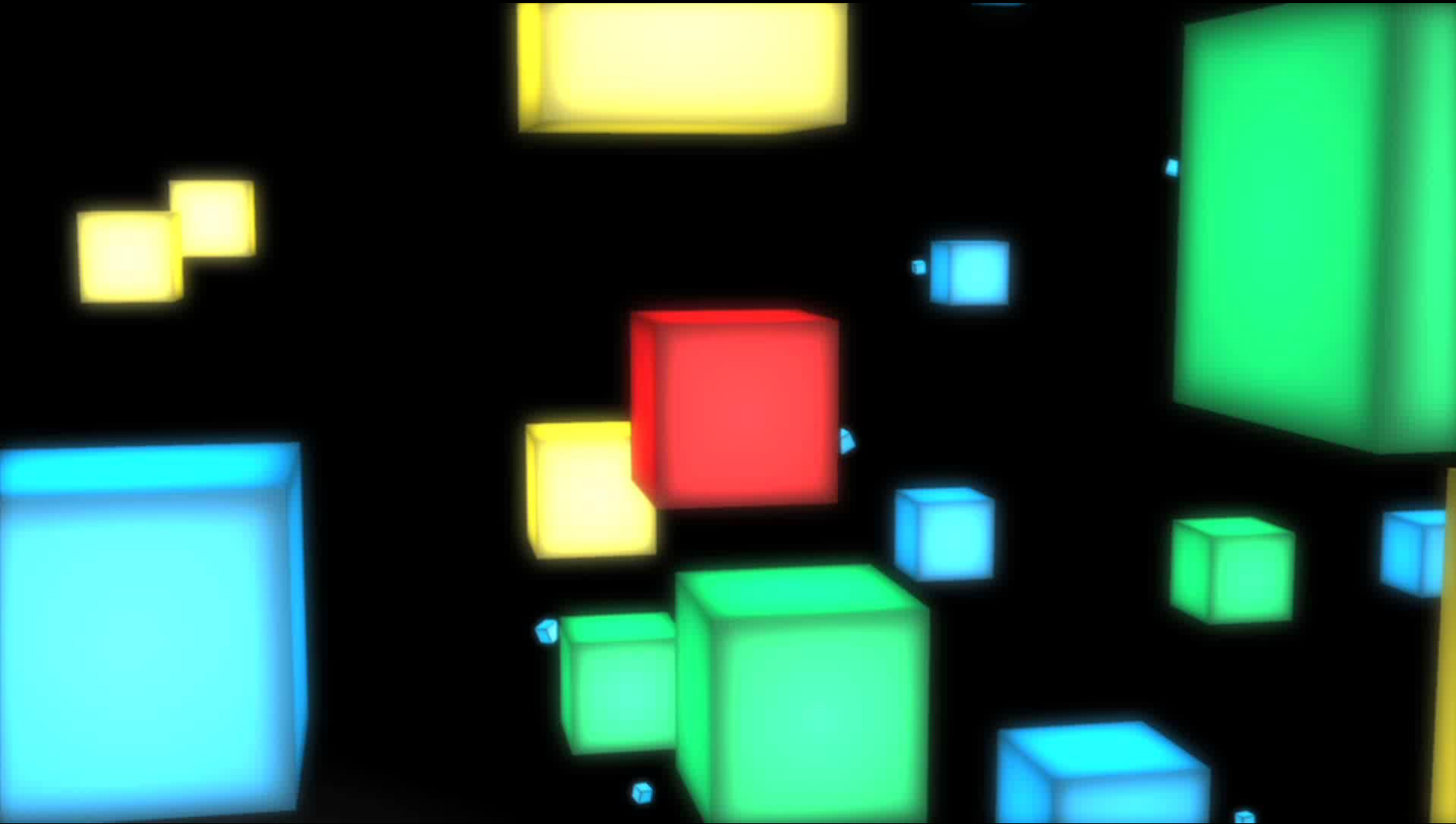
Modified and unmodified coding and non-coding RNA

Phosphorylated and unphosphorylated proteins

Metabolites

That promise to enable the construction of molecular networks that define the biological processes that comprise living systems

Integrating data to build predictive models of living systems
(■ - DNA, ■ - RNA, ■ - Protein, ■ - Metabolite)

An integrative genomics approach to the reconstruction of gene networks in segregating populations

J. Zhu,[a] P.Y. Lum,[a] J. Lamb,[a] D. GuhaThakurta,[a] S.W. Edwards,[a] R. Thieringer,[b] J.P. Berger,[c] M.S. Wu,[d] J. Thompson,[e] A.B. Sachs,[a] and E.E. Schadt[a]
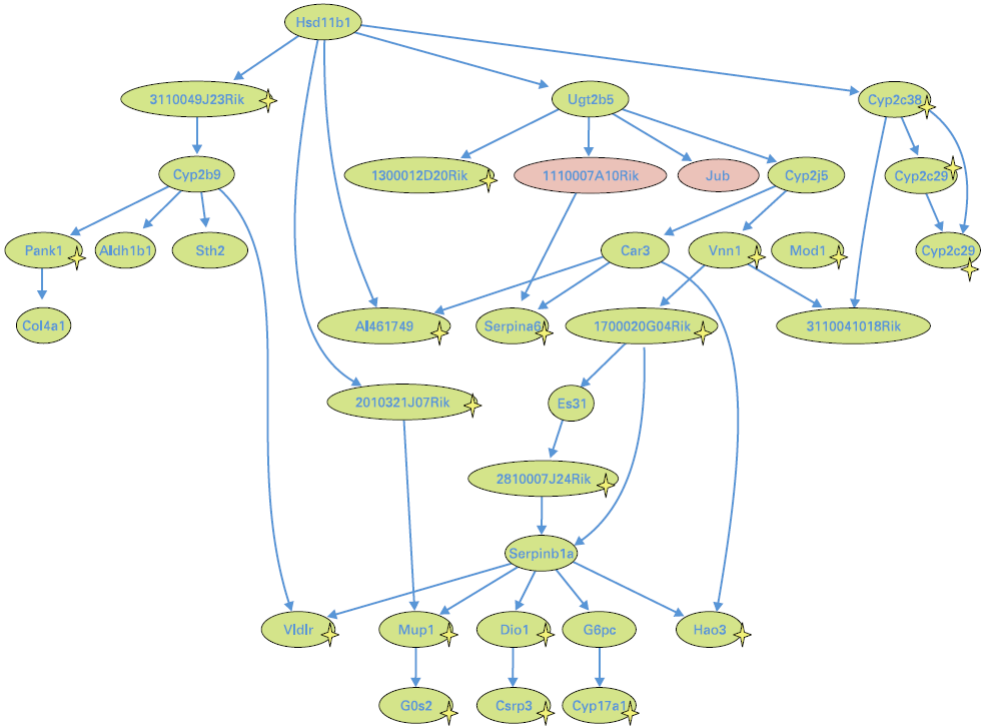
Leveraging DNA variations as a perturbation source is key to inferring causality
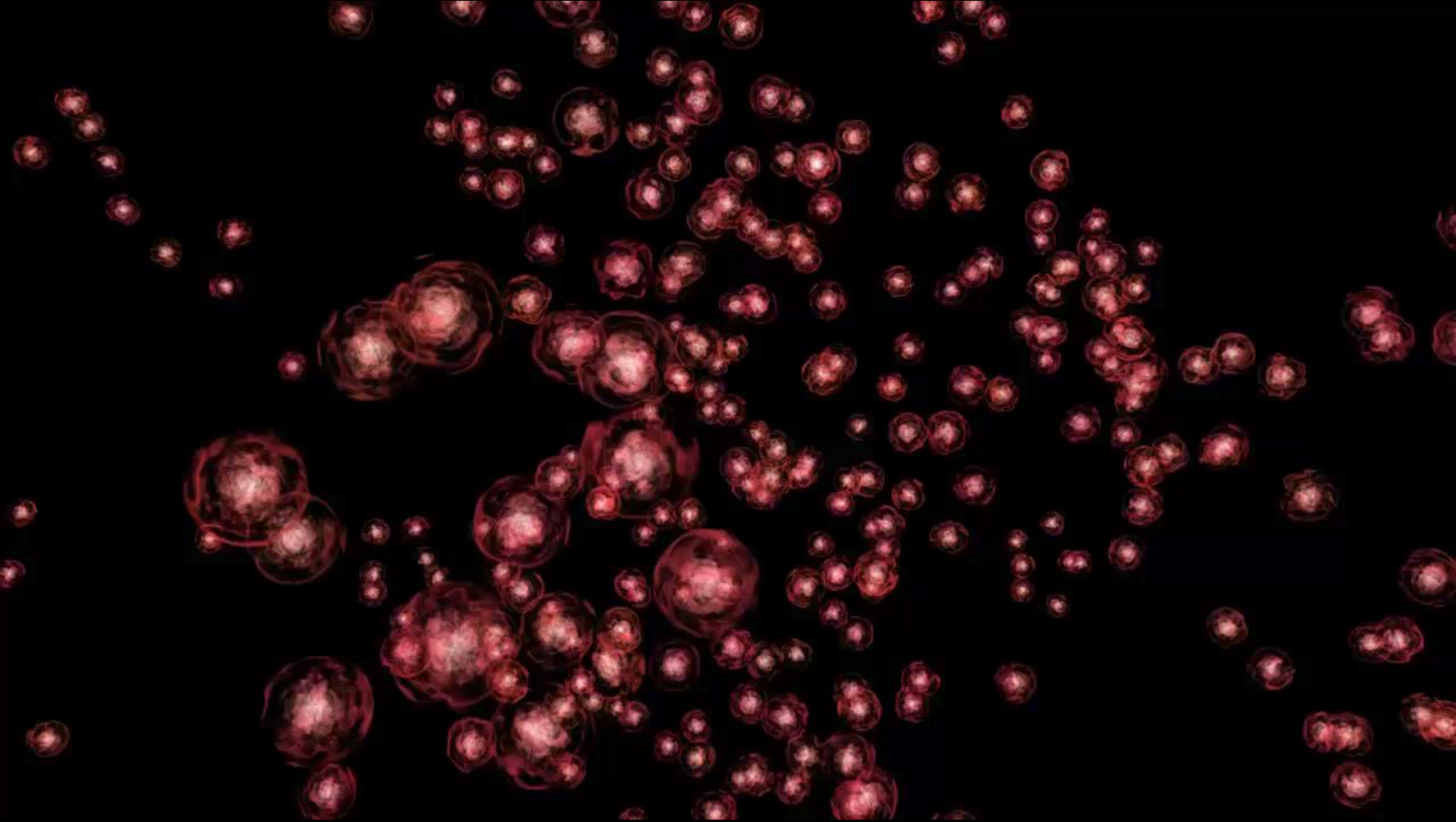
Understanding the network architecture critical for understanding how information flows through it
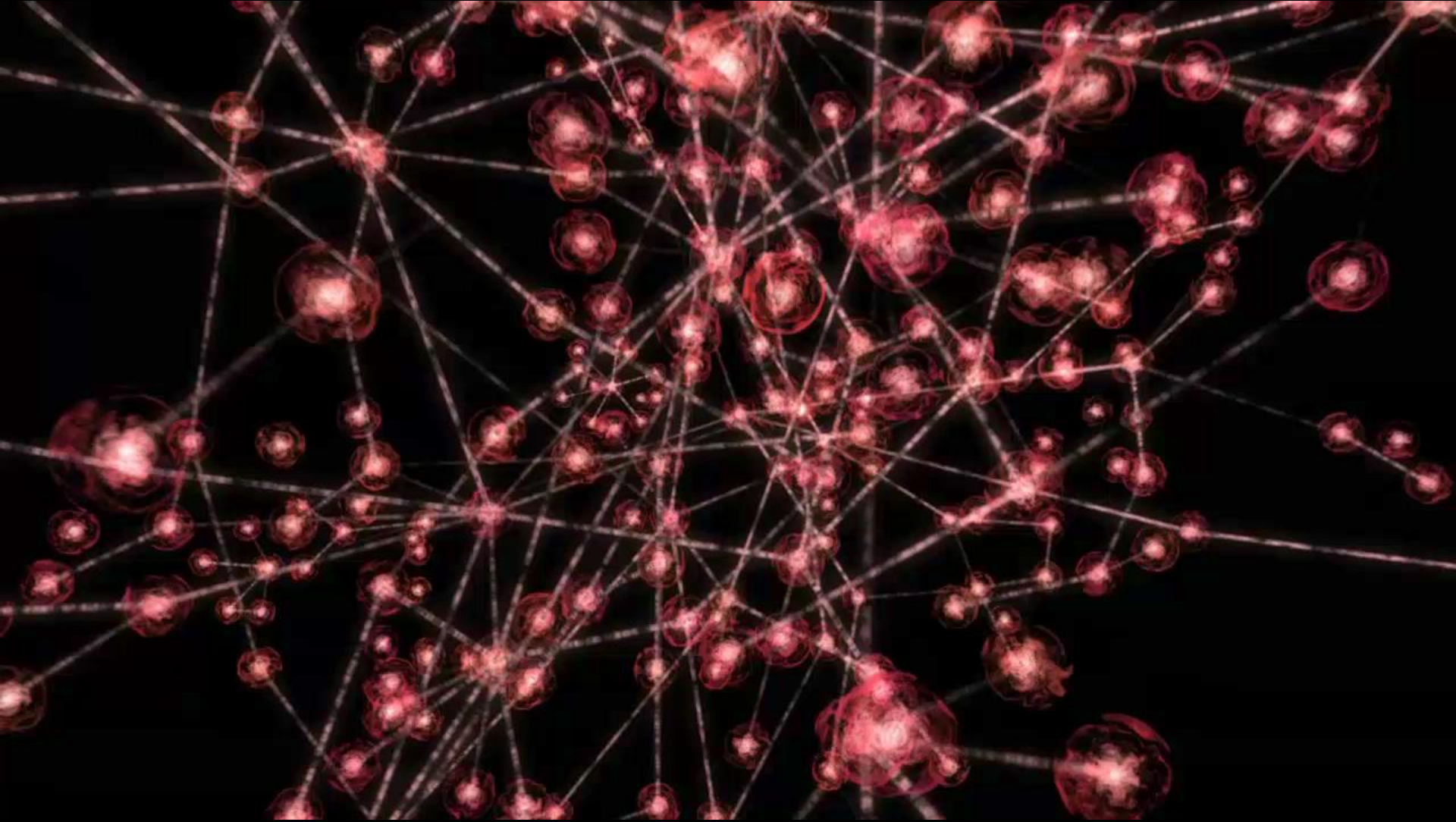
Stratifying patient populations

Input

Genome networks

DNA regulatory proteins

A

B

C

Molecular signaling

D

RNA networks

E

Protein network

Integrative network model of genes and traits
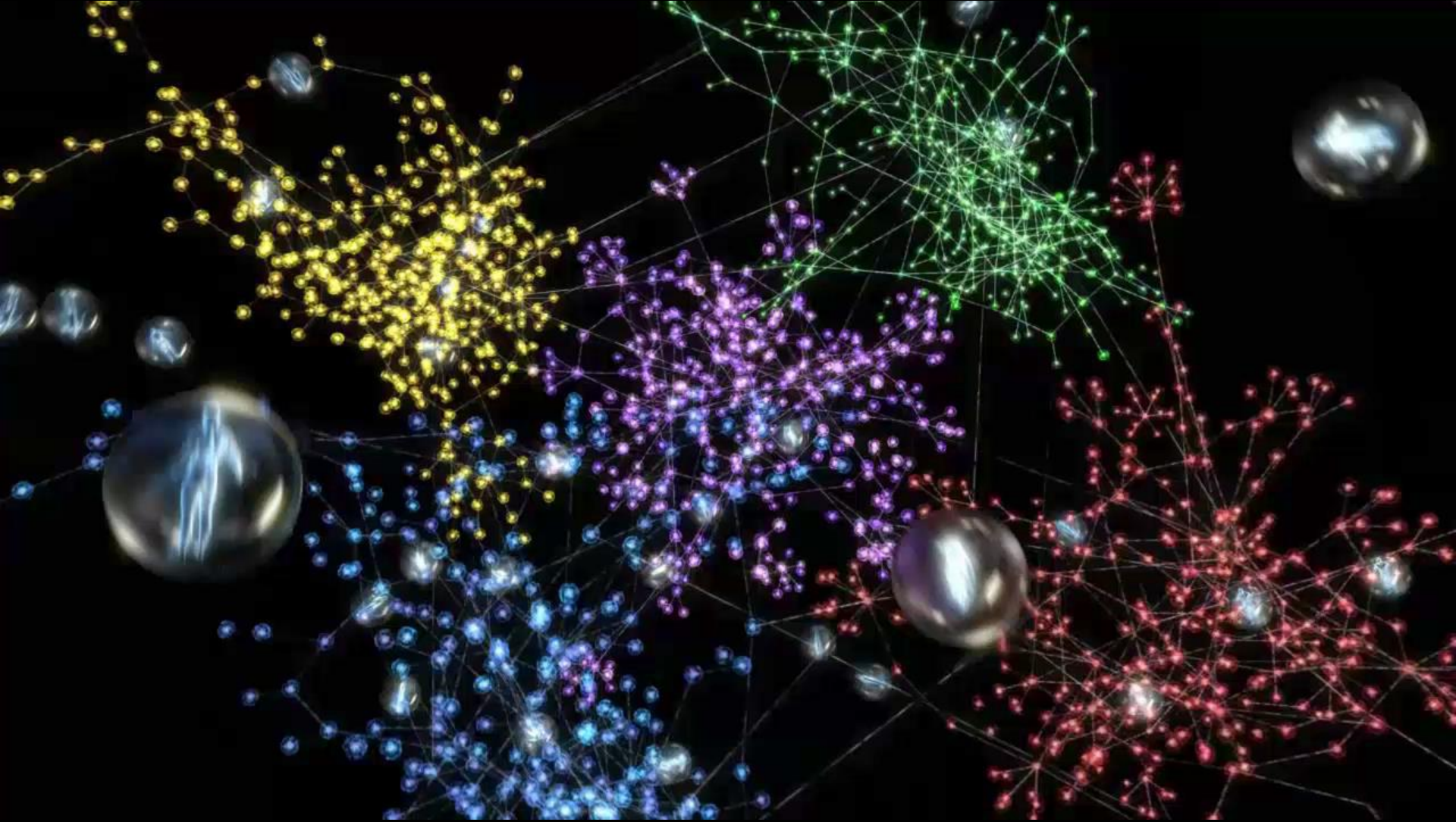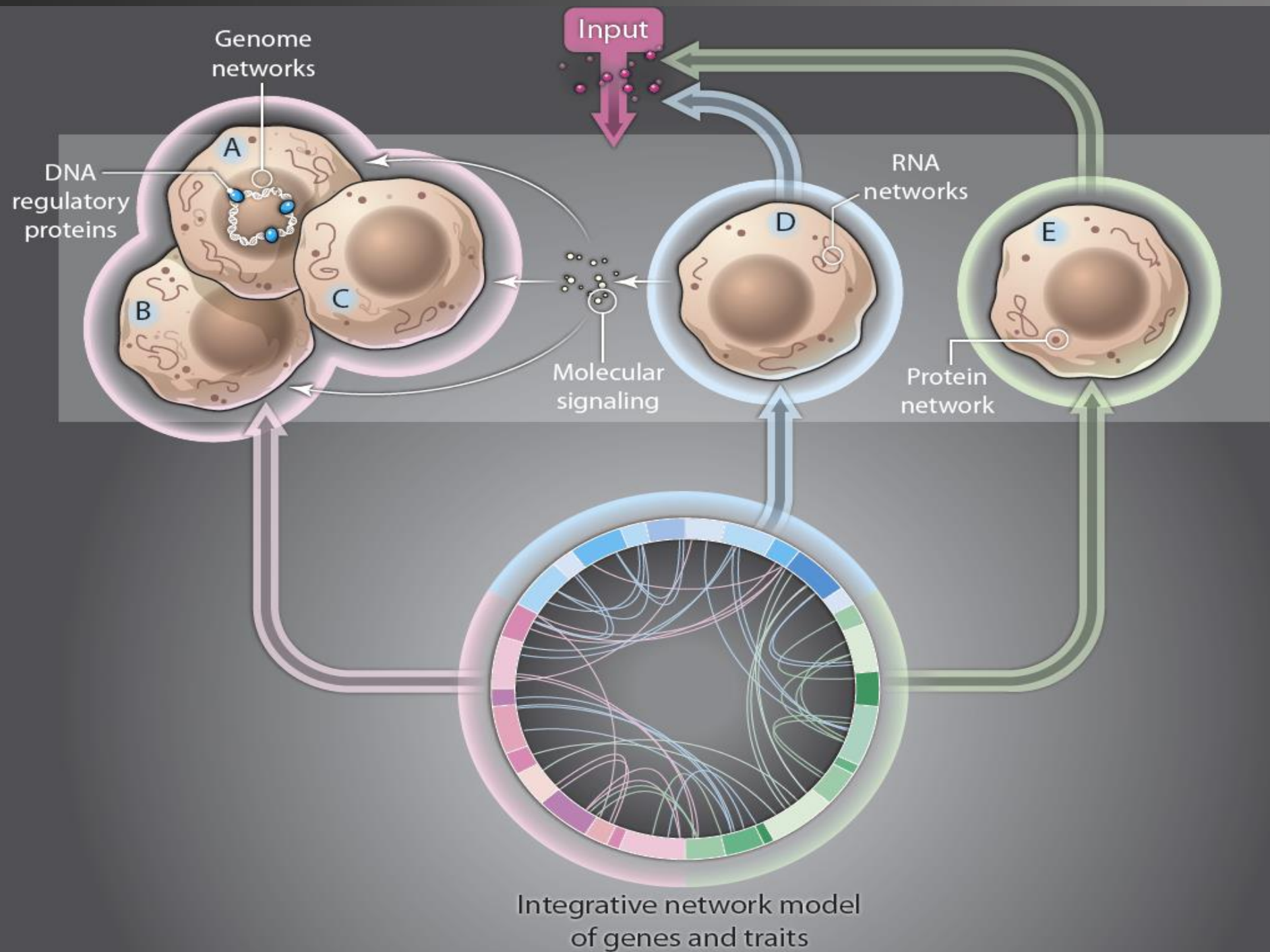
Organizing 163 genetic loci for IBD

nature

Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease

Problem: How do you make sense of 163 loci to understand a complex disease like IBD?

The Omental Adipose
Network

(Created with iCAVE from Gumus Lab, 2013)

(Created with iCAVE from Gumus Lab, 2013)

Building networks from 500 prefrontal cortex samples
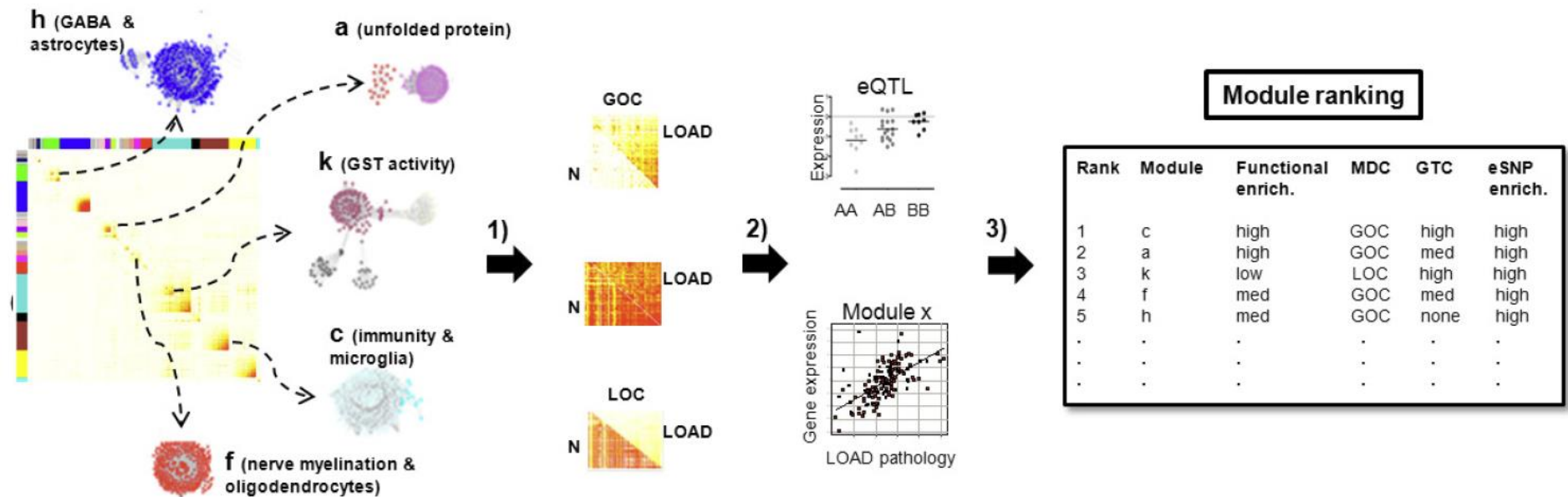
Constructing the co-expression networks
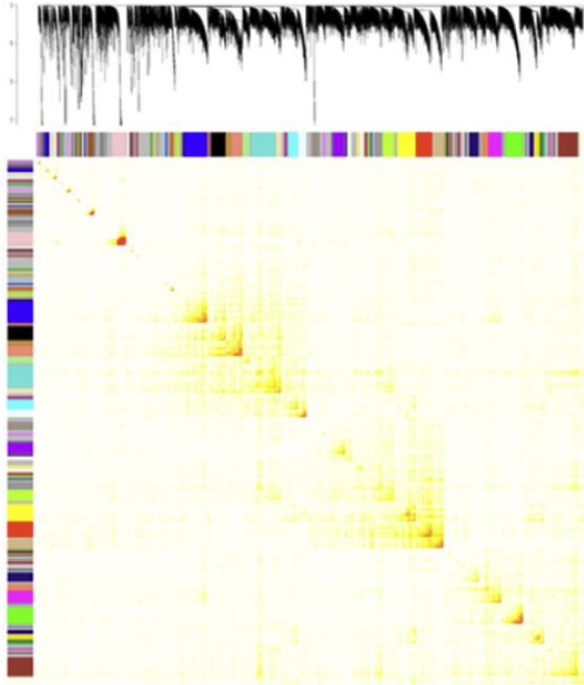
"Normal" versus LOAD Networks

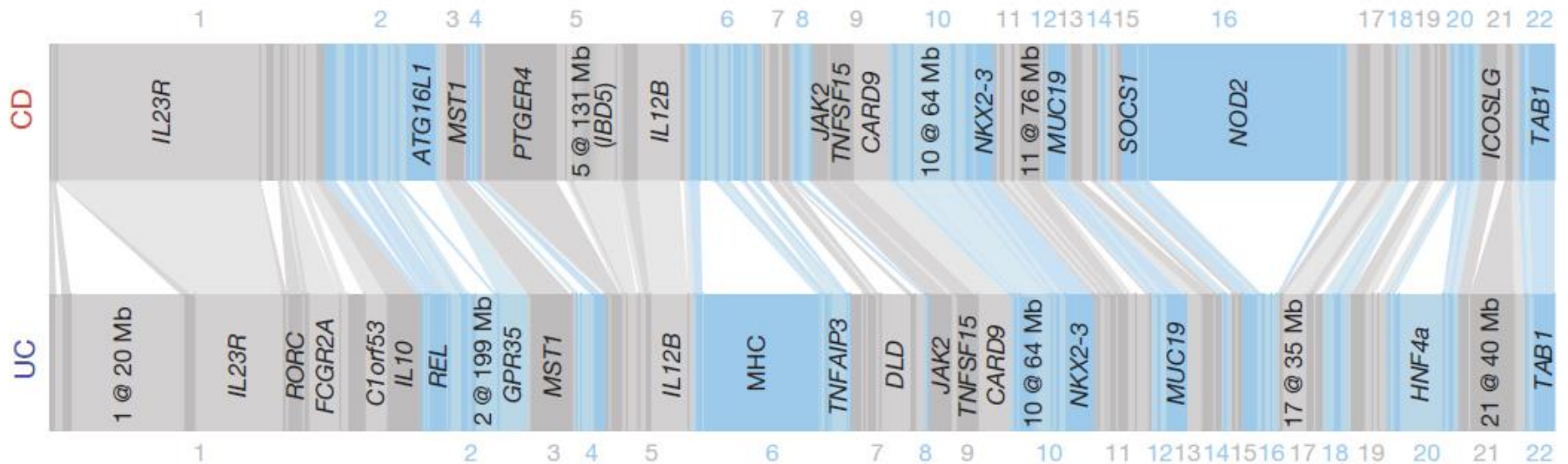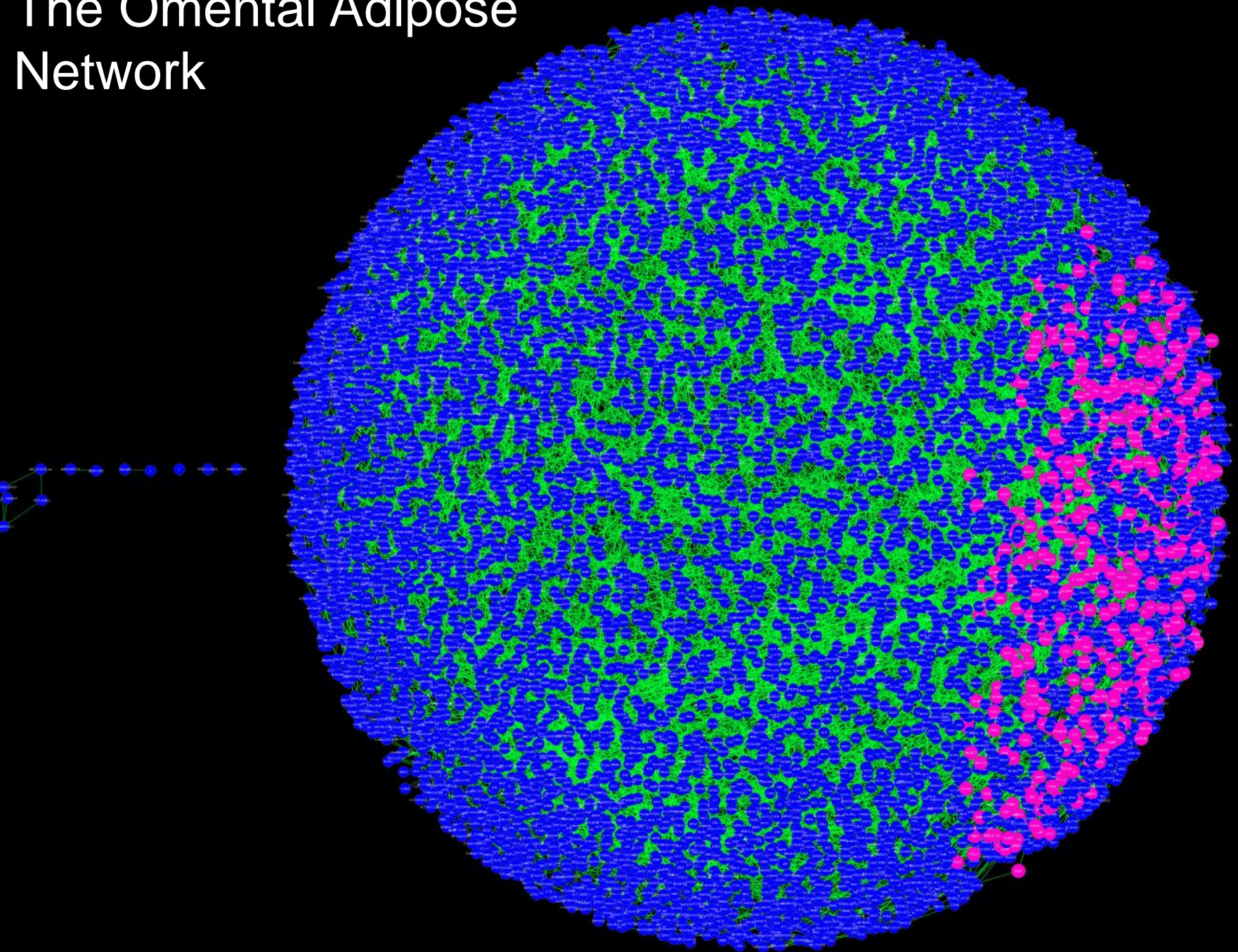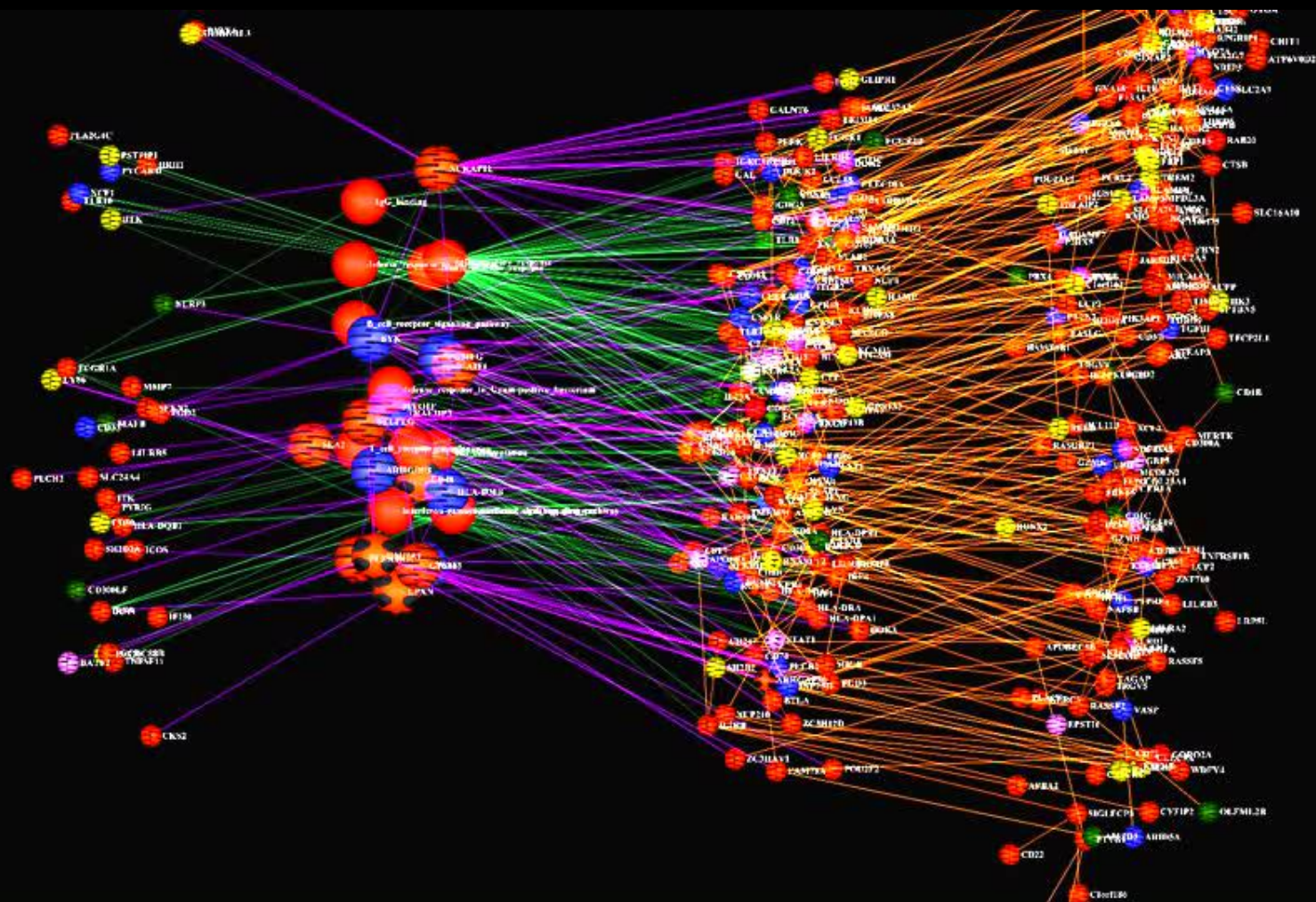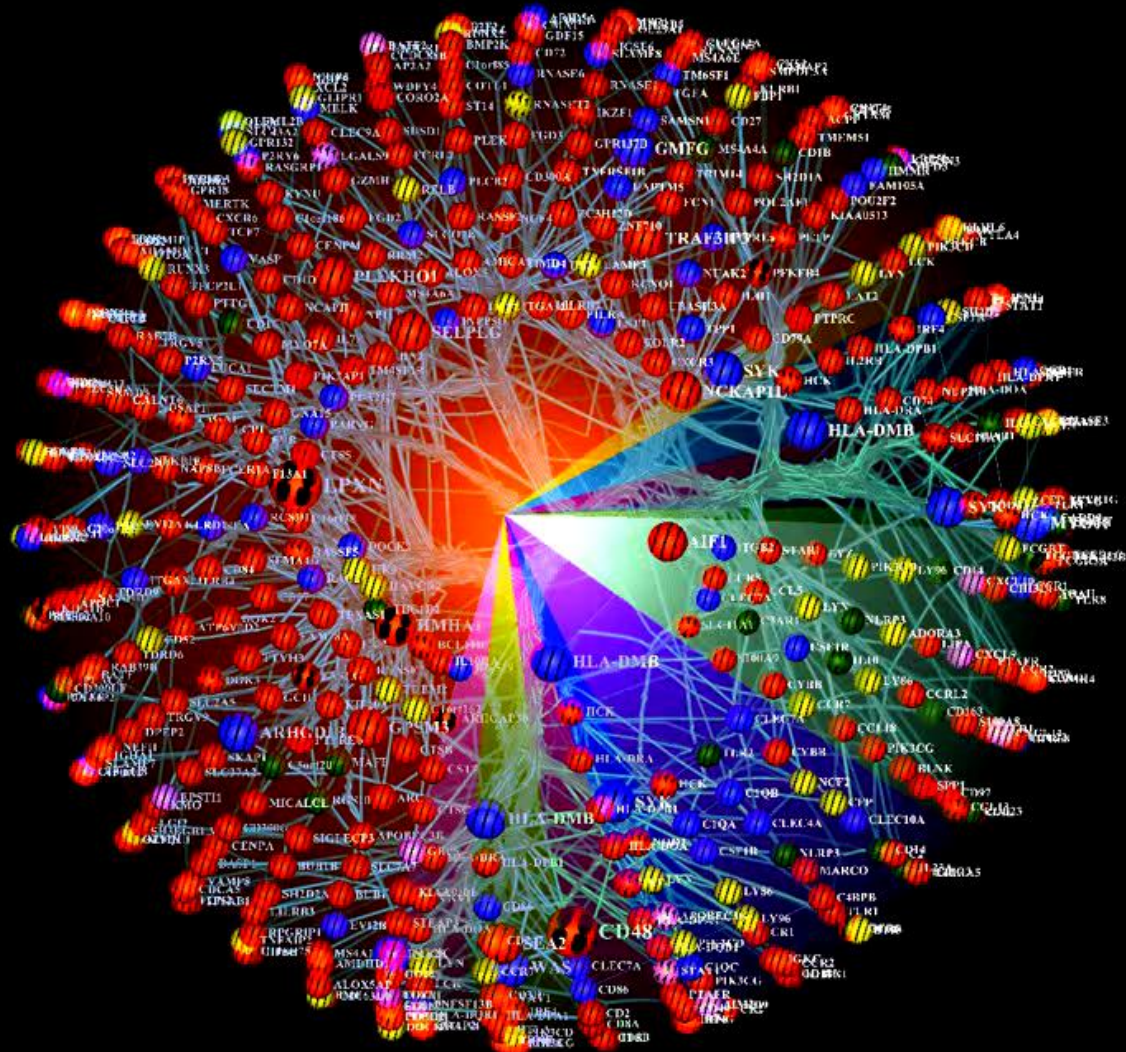| Pink GST activity (92.3, <0.001) LOAD Normal	Green yellow Unfolded protein (4.2, <0.001)	Light green Nerve myelination (0.2, <0.001)	Chartreuse Gluconyl transferase (2.3, 0.02)
Khaki GABA biosynthesis (0.3, <0.001)	Orchid Phototransduction (77.6, <0.001)	Chocolate Dynein complex (2.1, 0.002)	Red 3 Ribosome (24.9, <0.001)
Khaki GABA biosynthesis (0.3, 0.004)	Green 4 Cell cycle (0.3, <0.001)	Midnight blue Neurogenesis (8.0, <0.001)	Yellow Immune functions (2.0, <0.001)
Cyan 2 Muscle contraction (0.9, 0.23)	Brown 2 Olfactory perception (25.5, <0.001)	Firebrick Neutotrophin signaling (0.3, <0.001)	Tan Extracellular matrix (2.9, <0.001)

Causal probabilistic network relating to a PFC module correlating with multiple LOAD clinical covariates, enriched for immune function/pathways related to microglia activity

We identified TYROBP as a key regulator of this network

CD33, MS4A4A, MS4A6A (from LOAD GWAS)

Two papers in NEJM today reporting on rare variants in TREM2 associate with LOAD

molecular systems biology

# Systems analysis of eleven rodent disease models reveals an inflammatome signature and key drivers
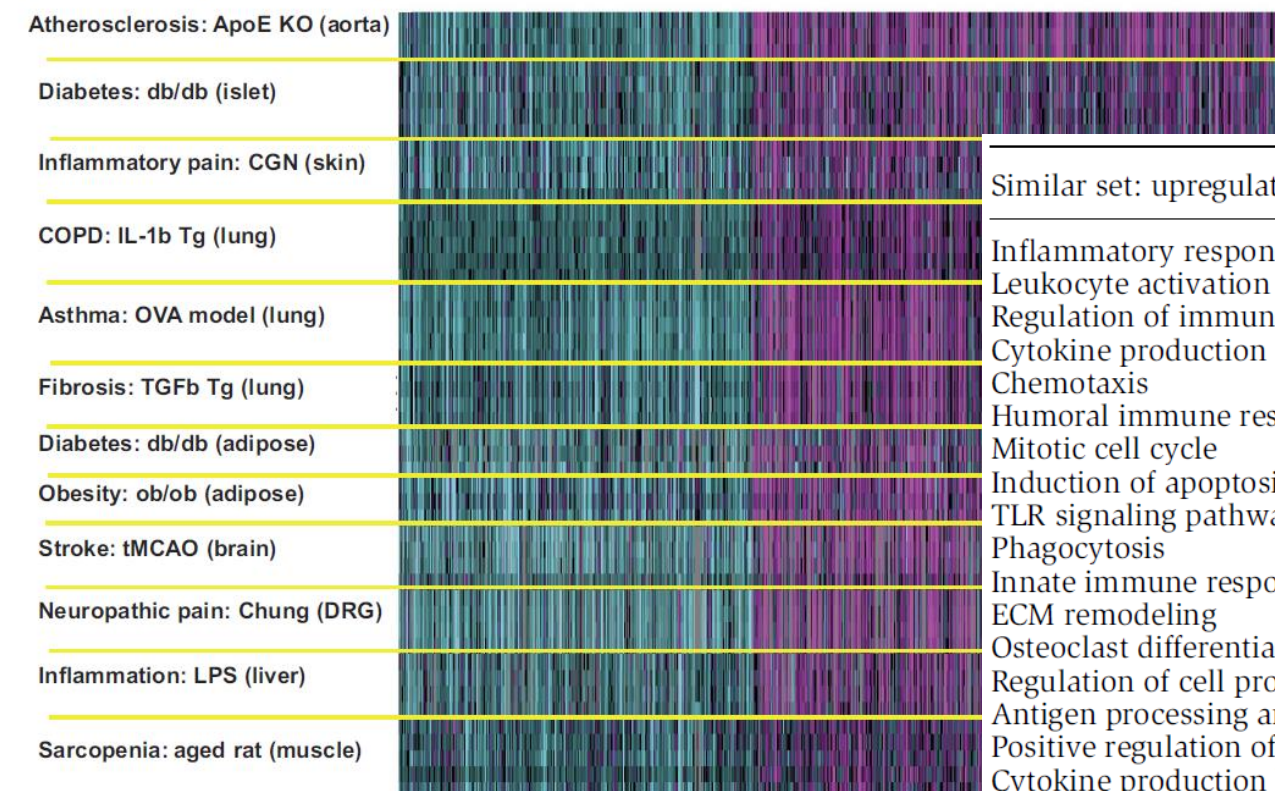
Similar set: upregulated	Enrichment P	Overlap	Set
Inflammatory response	4.76E − 61	208	704
Leukocyte activation	2.13E − 32	164	704
Regulation of immune response	1.44E − 25	84	260
Cytokine production	6.10E − 18	85	335
Chemotaxis	4.97E − 16	74	284
Humoral immune response	3.25E − 14	69	271
Mitotic cell cycle	7.64E − 13	87	414
Induction of apoptosis	1.74E − 12	86	412
TLR signaling pathway	4.66E − 12	21	47
Phagocytosis	2.74E − 11	38	111
Innate immune response	9.29E − 11	48	173
ECM remodeling	9.67E − 11	22	59
Osteoclast differentiation	3.61E − 10	31	82
Regulation of cell proliferation	4.19E − 10	112	662
Antigen processing and presentation	7.44E − 10	32	89
Positive regulation of translation	3.64E − 09	45	170
Cytokine production by Th17 cells	6.33E − 09	17	40
Angiogenesis	9.41E − 09	60	277
Cell-cycle process	2.79E − 08	104	636
Wound healing	1.79E − 07	57	274
Regulation of translation	4.44E − 07	66	349
Macrophage activation	1.61E − 06	20	49
Interleukin-12 production	1.70E − 06	18	40

# Core disease modules harbor pluripotent drug targets
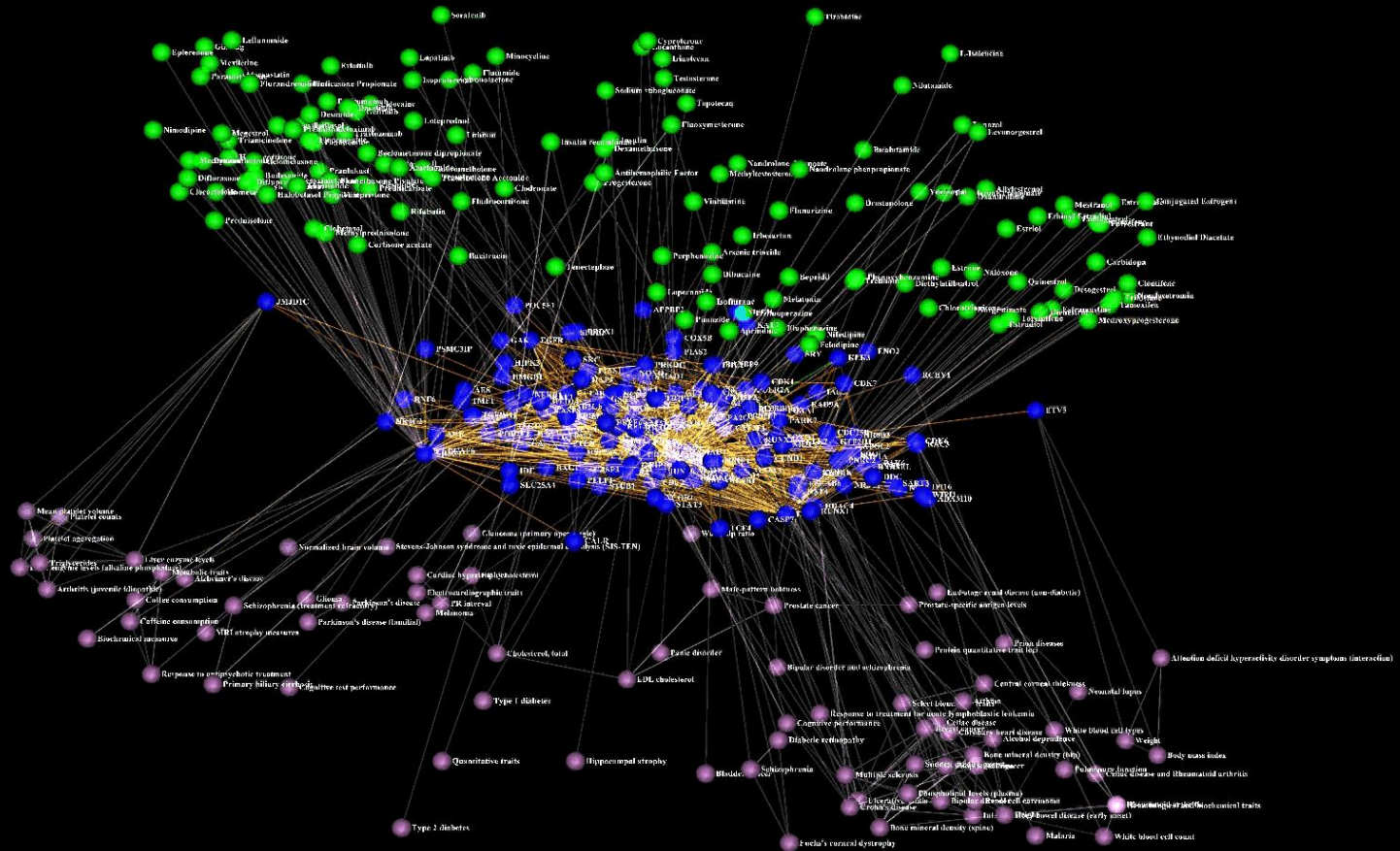
Functional chemigenomics screen: Chemical perturbagens against disease networks *in silico*
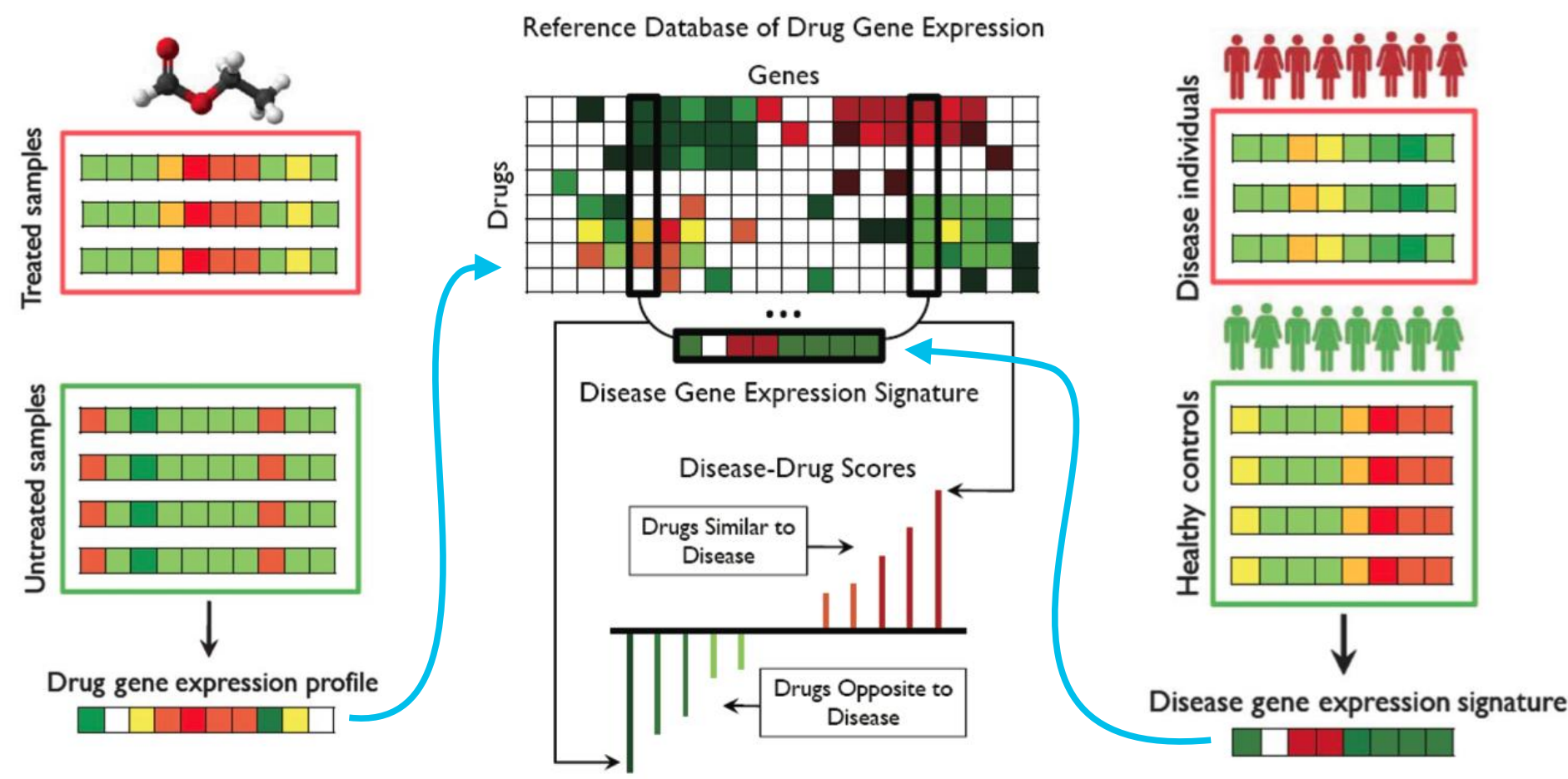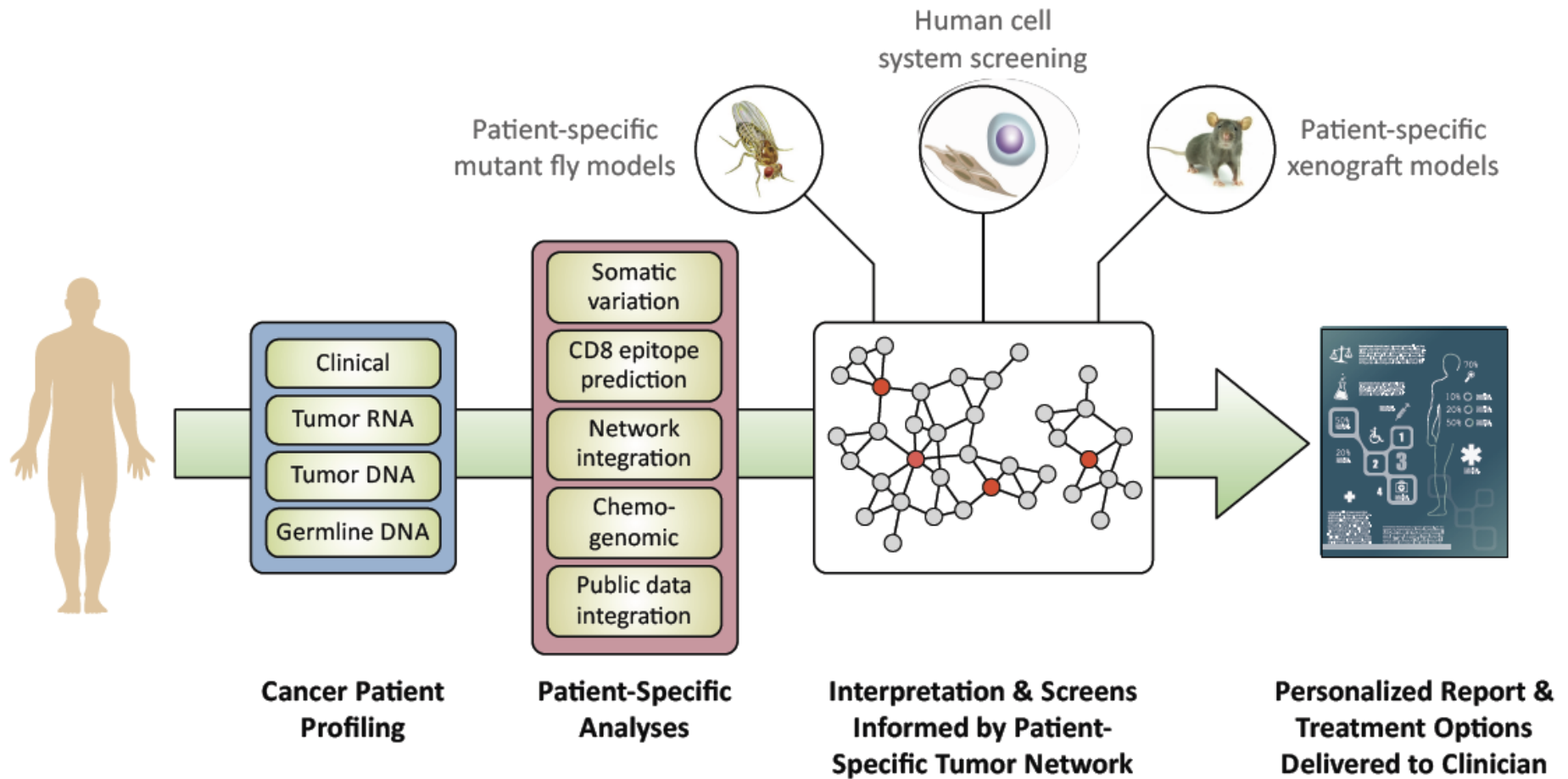
Topiramate Reduces IBD Severity in a TNBS Rodent Model of IBD

- TNBS chemically induced rat model of IBD

- Animals treated with 80mg/kg topiramate oral after sensitization

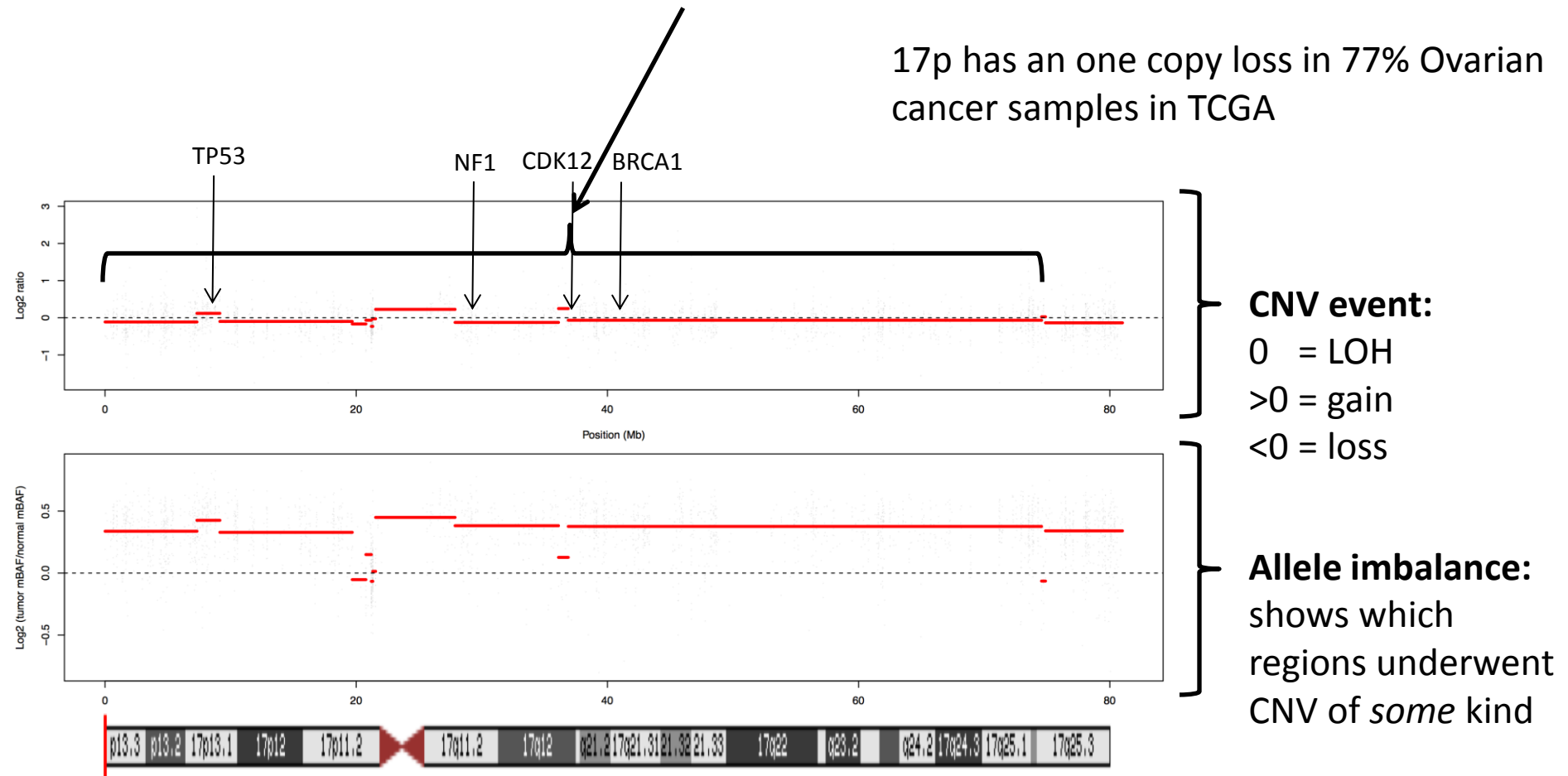- Prednisolone positive control (approved for IBD in humans)
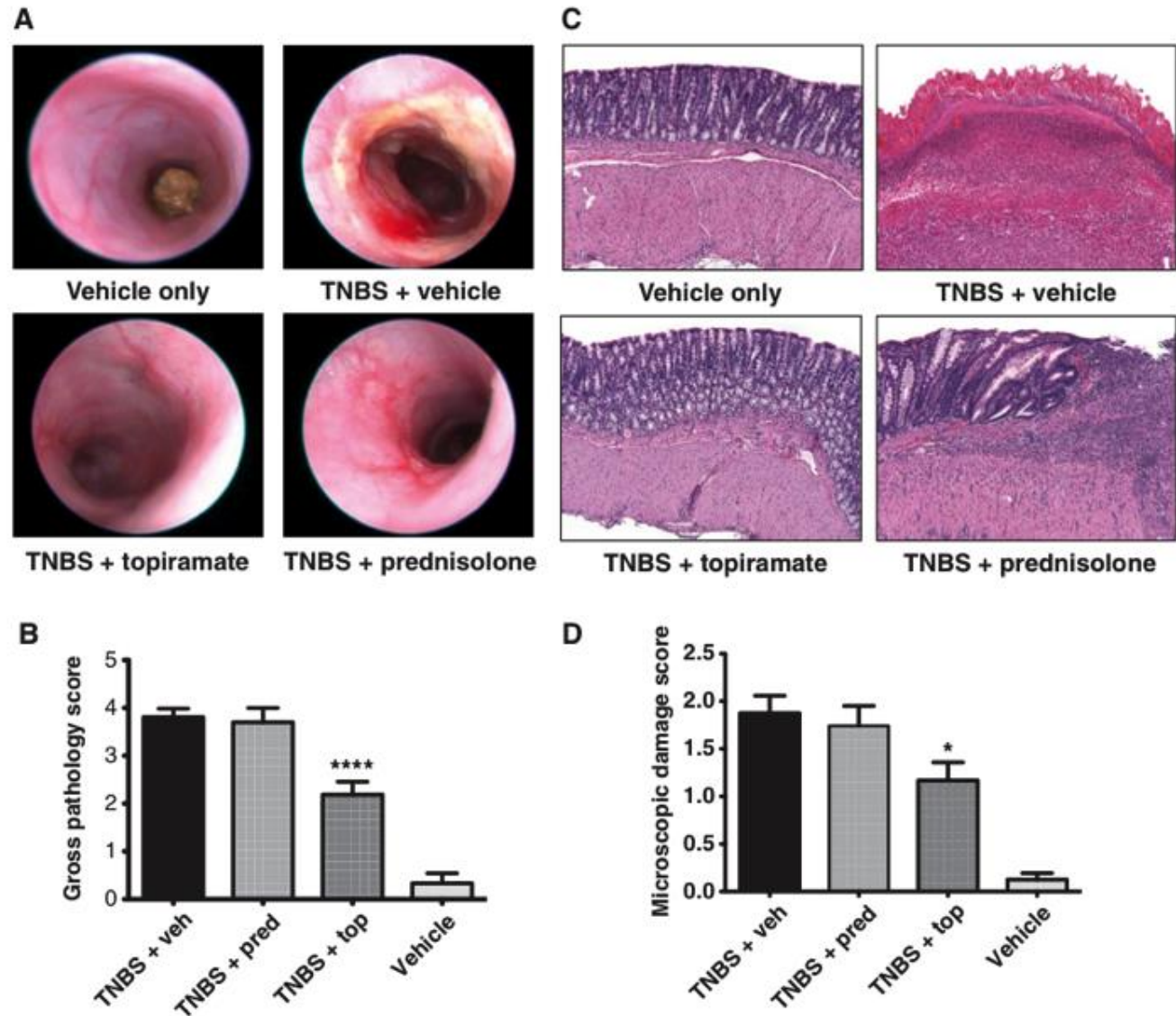
Leveraging NGS and Predictive Network Models to Drive Personalized Cancer Therapy
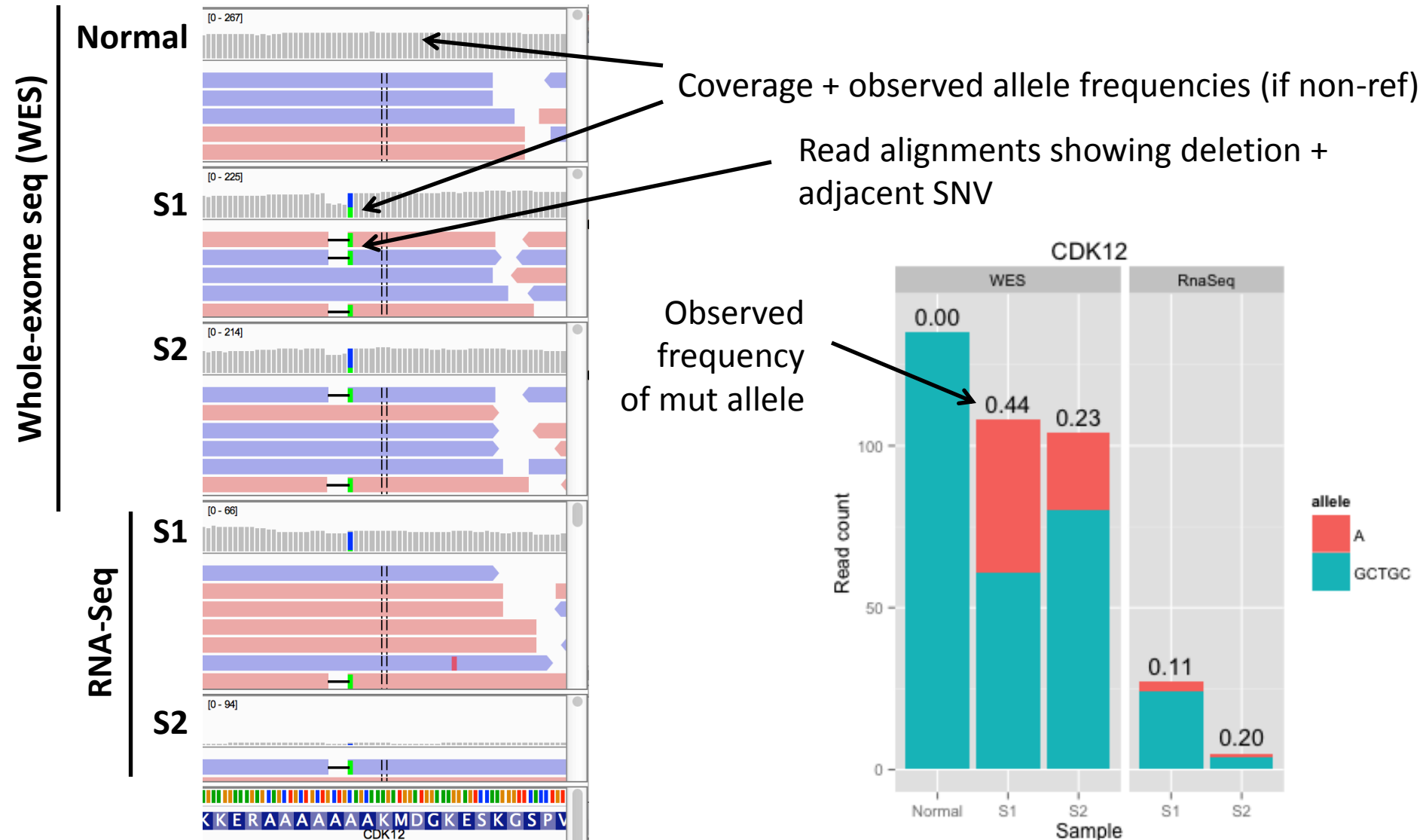
chr17 in tumor S1 underwent somatic copy number loss

LOH of the whole chromosome 17, which includes
TP53, **BRCA1**, **CDK12**, **ERBB2**, **TRIM37**
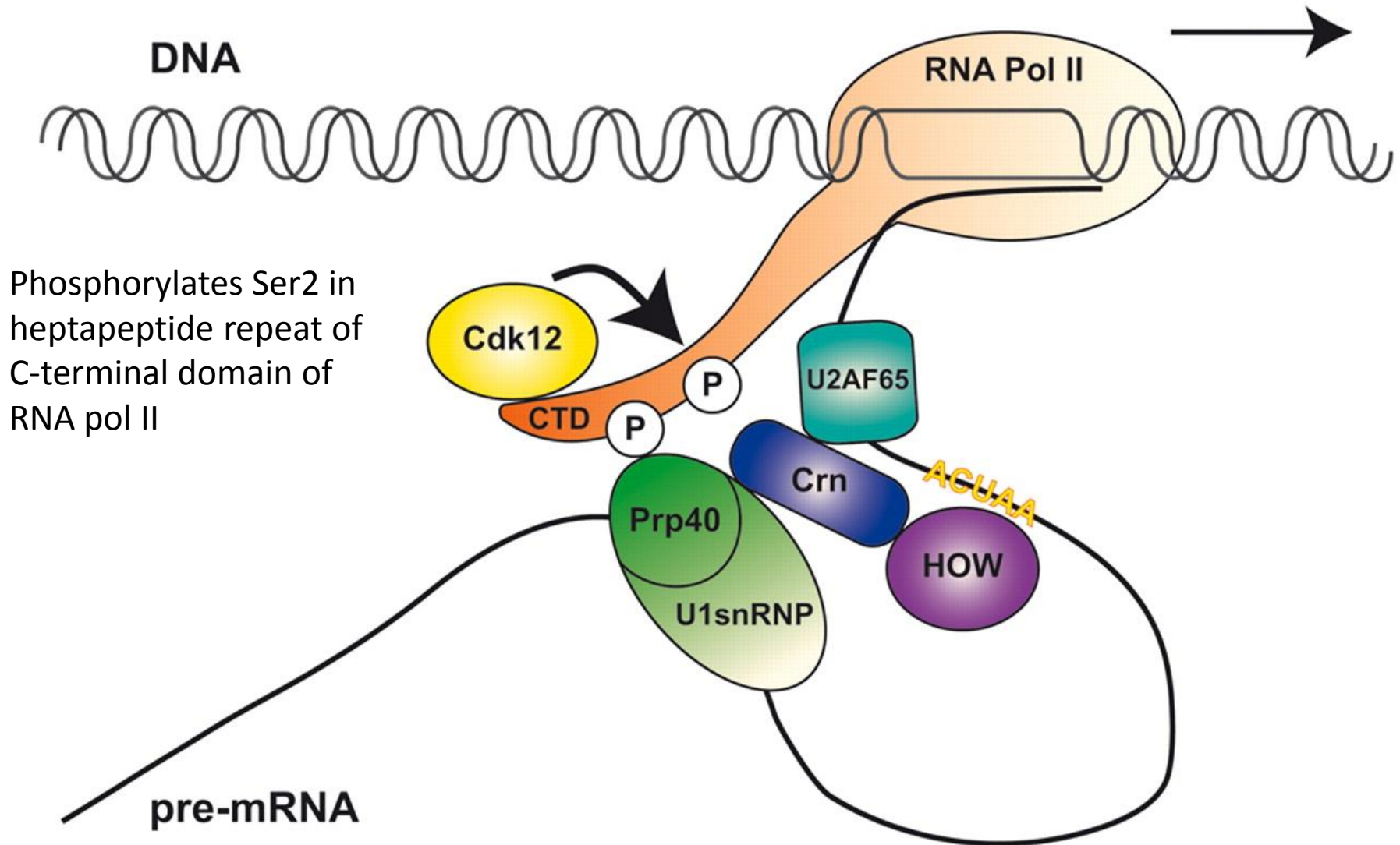
17p has an one copy loss in 77% Ovarian cancer samples in TCGA

CNV event:
0   = LOH
>0 = gain
<0 = loss
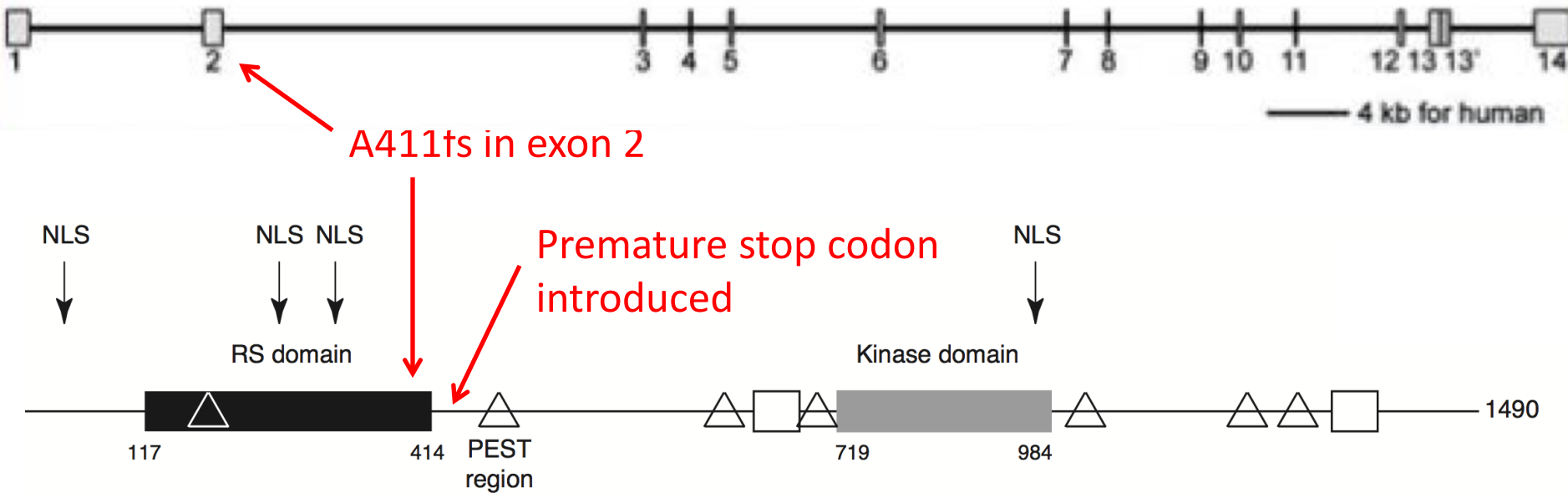
Allele imbalance:
shows which regions underwent CNV of *some* kind

Frameshift deletion A411fs found in CDK12 in both sites

CDK12 primes HOW/Crn-dependent splicin in fly glial cells

DNA

RNA Pol II

Phosphorylates Ser2 in heptapeptide repeat of C-terminal domain of RNA pol II

Cdk12

CTD

P

P

U2AF65

Crn

ACUAA

Prp40

HOW

U1snRNP

pre-mRNA

CDK12 mutation results in loss of kinase domain

Normal: aa401 RKKKERAAAAAAAKMDGKESKGSPVFLPRKENSSVEAKDS...
Mutant: aa401 RKKKERAAAAKQRWMERSPRVHLYFCLEKRTVQ*

NLS: nuclear localization signal
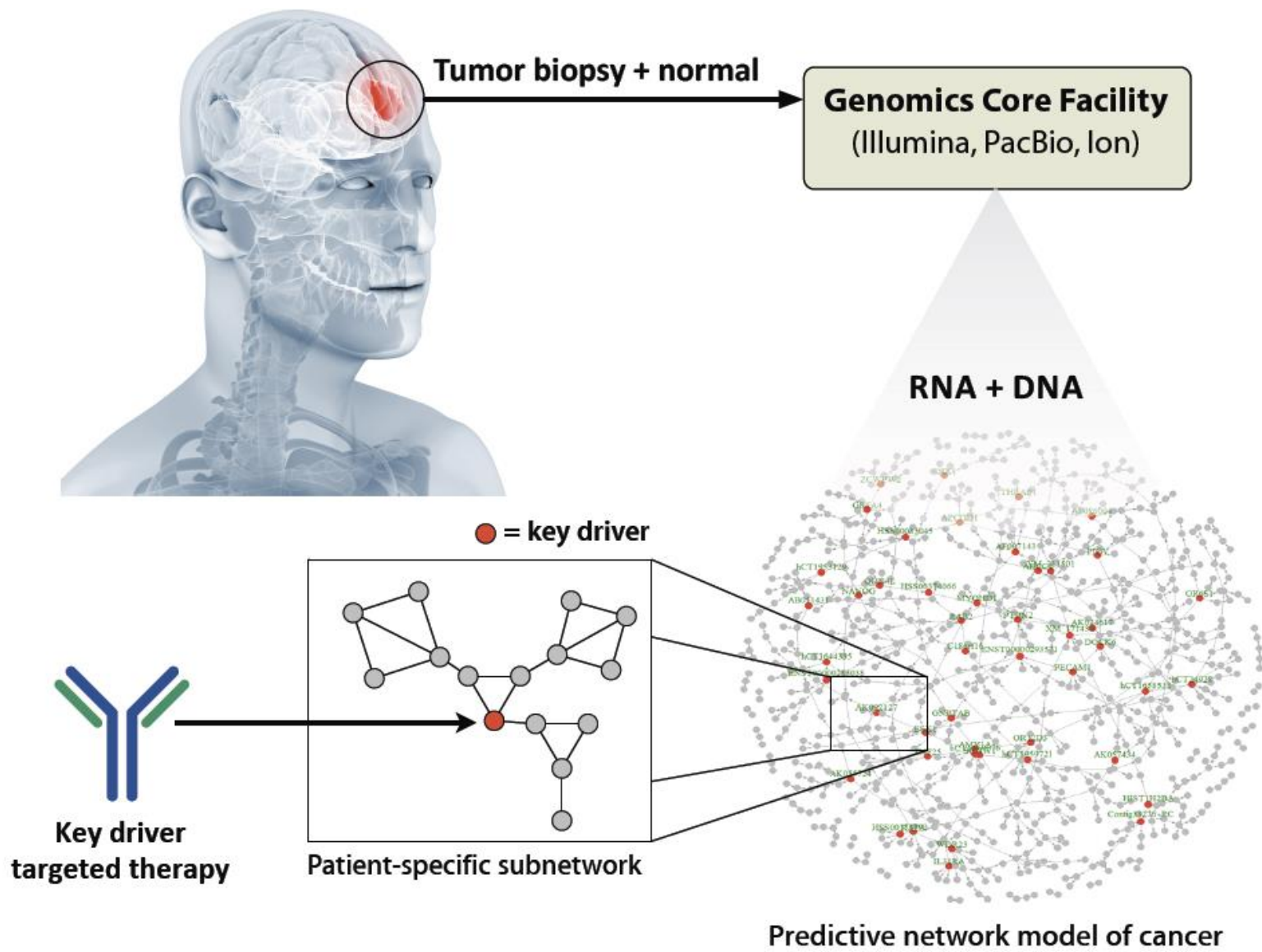**RS domain:** arginine/serine-rich domain
PEST region: peptide sequence rich in proline
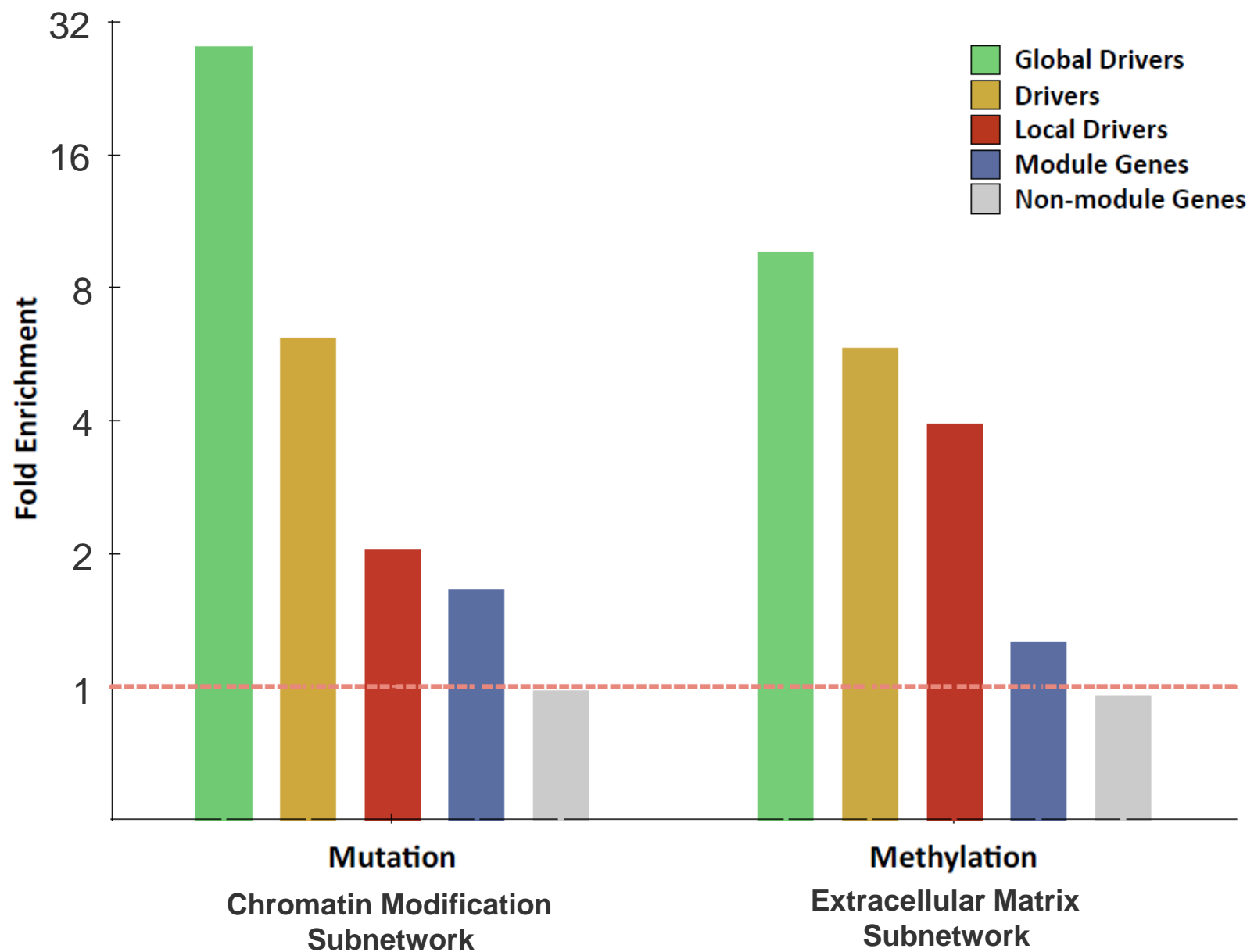**kinase domain:** serine-threonine kinase domain
PRM: proline-rich motif

Ko TK (J Cell Sci 2001)
Chen HH (Mol Cel Biol 2006)
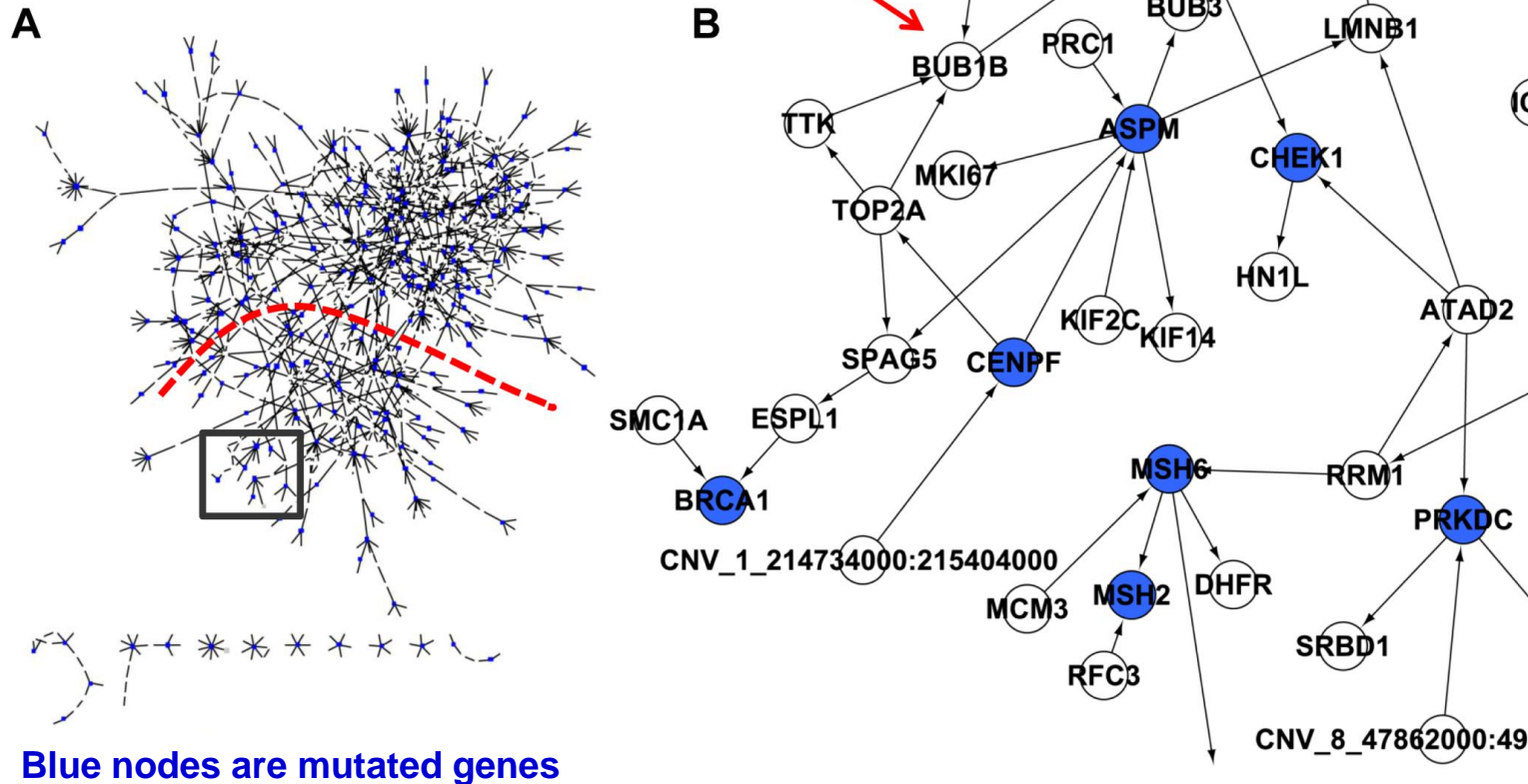Taglialatela A (PhD thesis 2012)

Personalized multiscale tumor networks to diagnose and treat cancers

Key driver analysis: Identifying those genes that regulate network states that have larger impact on outcomes

Patient mutation data projected onto the network: Interesting 1000-node subnetwork identified


Blue nodes are mutated genes

- Full network comprised of 7,881 expr/2,331 CNV nodes, 306 regulators, 501 functional mutations
- Subnetwork: 116 regulators, 232 functional mutations – massive enrichment (p = 4.5e-173)
- 6 mutations affecting master regulators in patient and TCGA data, including ASPM and CENPF related to BUB1B dependency
- Many pathways dramatically enriched: transmembrane receptor protein tyrosine kinase activity, collagen binding, axonogenesis and so on

Using multiscale tumor networks to inform personalized chemotherapy options
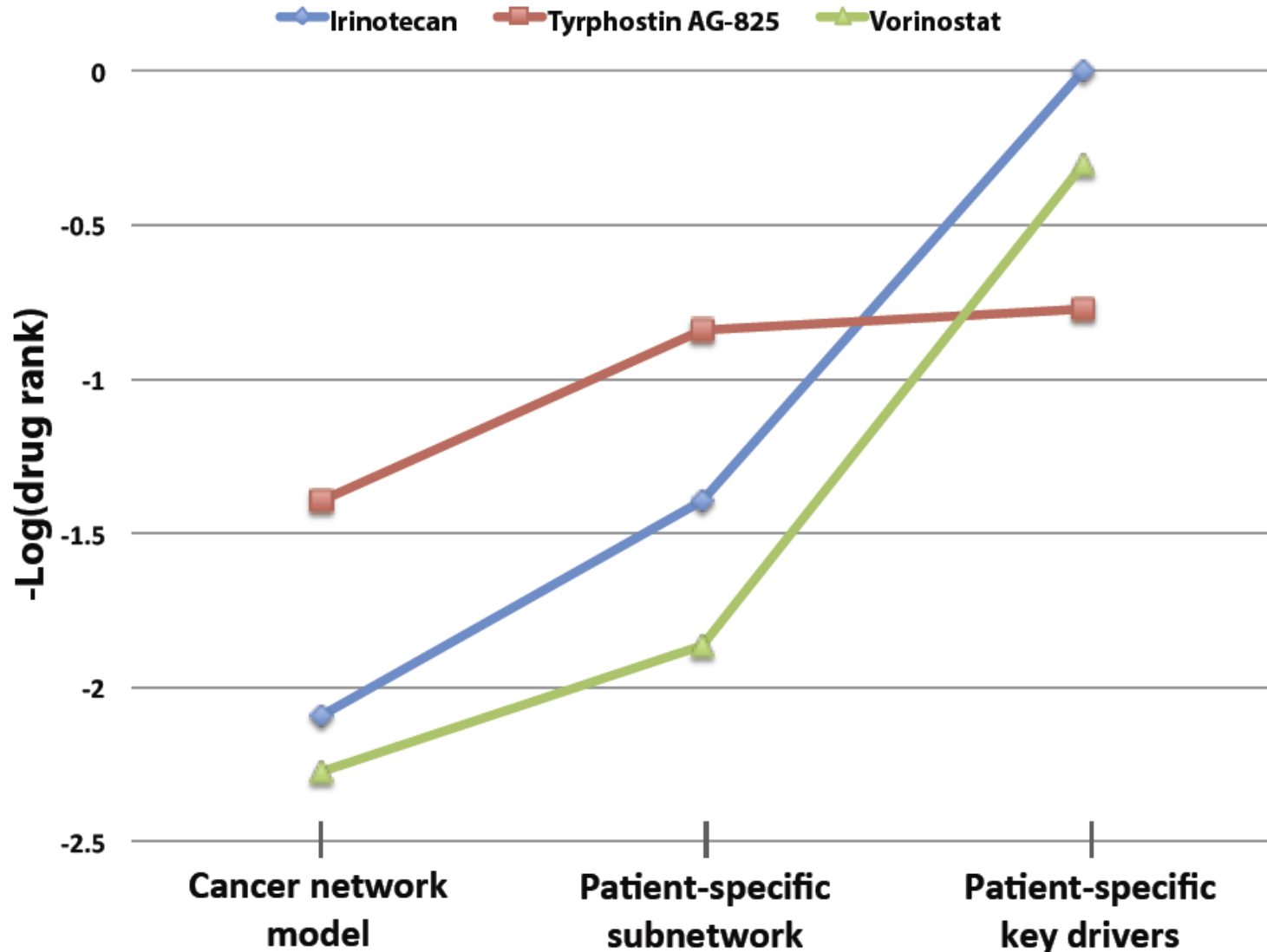
Aiming to build personalized multiscale networks to model dynamics of complex disease

DNA
Cell-specific RNA
Cytokines
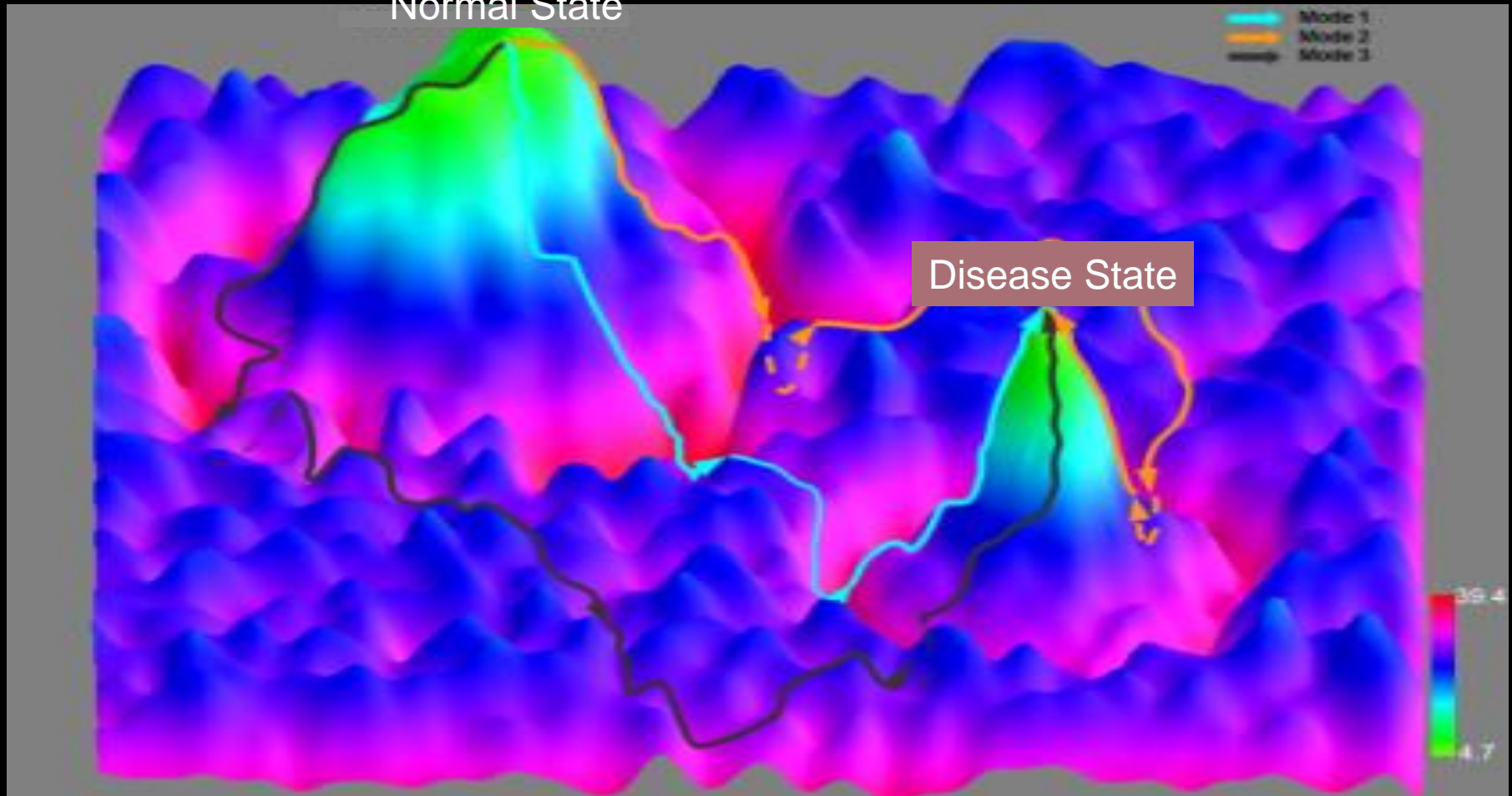Clinical labs
Physiometrics

0:00 min

0:05 min

0:10 min
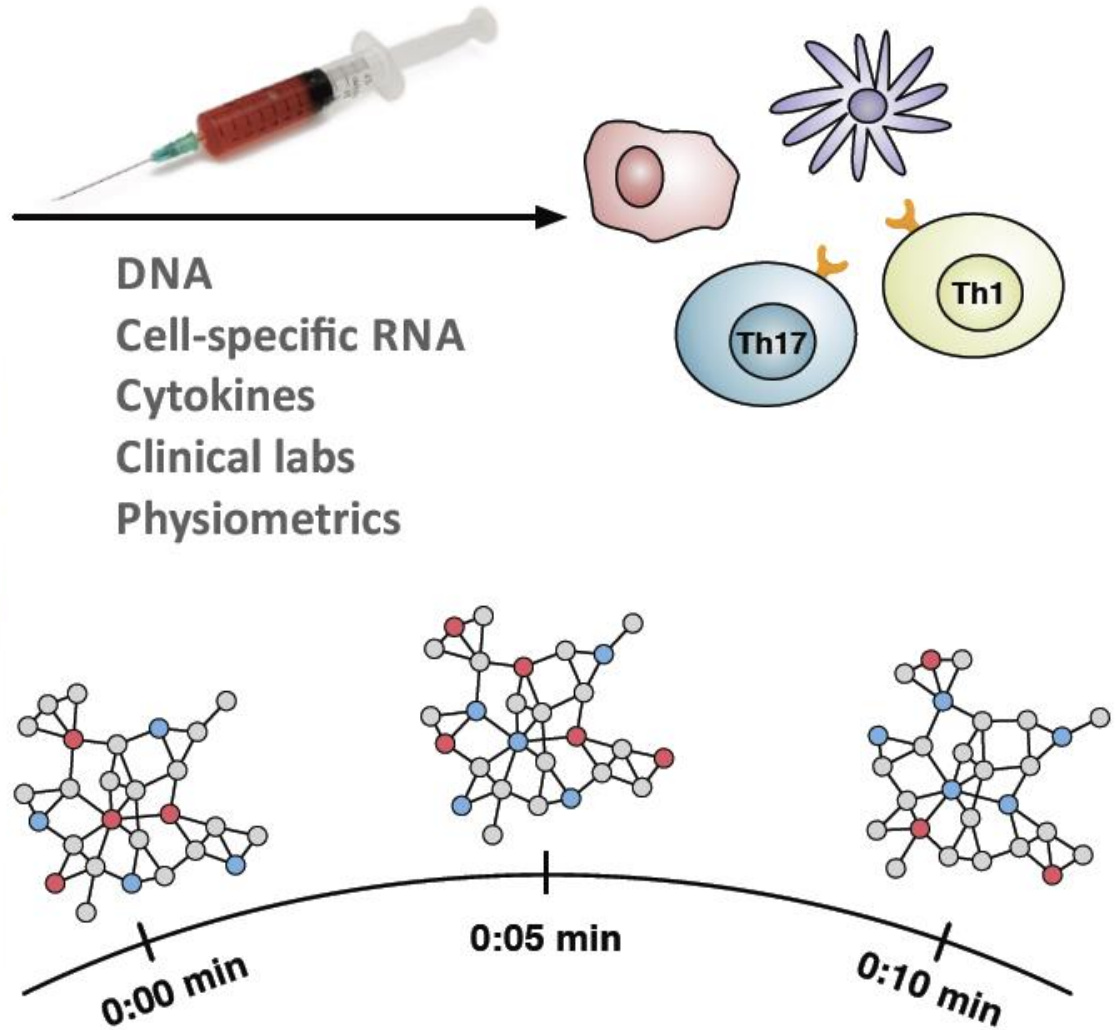
High-dimensional data acquisition carried out over time and at multiple scales can provide for the precision medicine approach we all seek

● Genomic
● Environment
○ Clinical
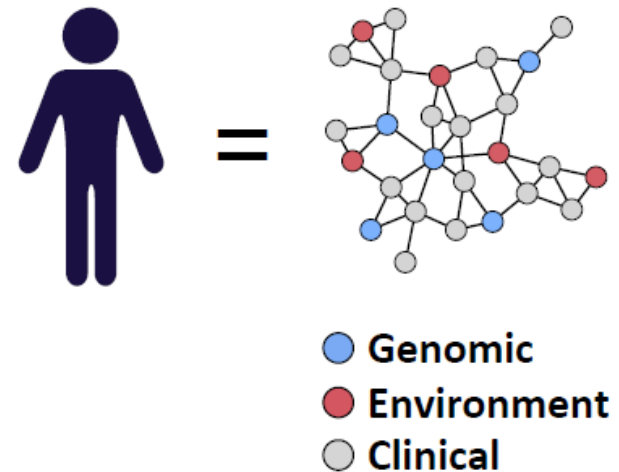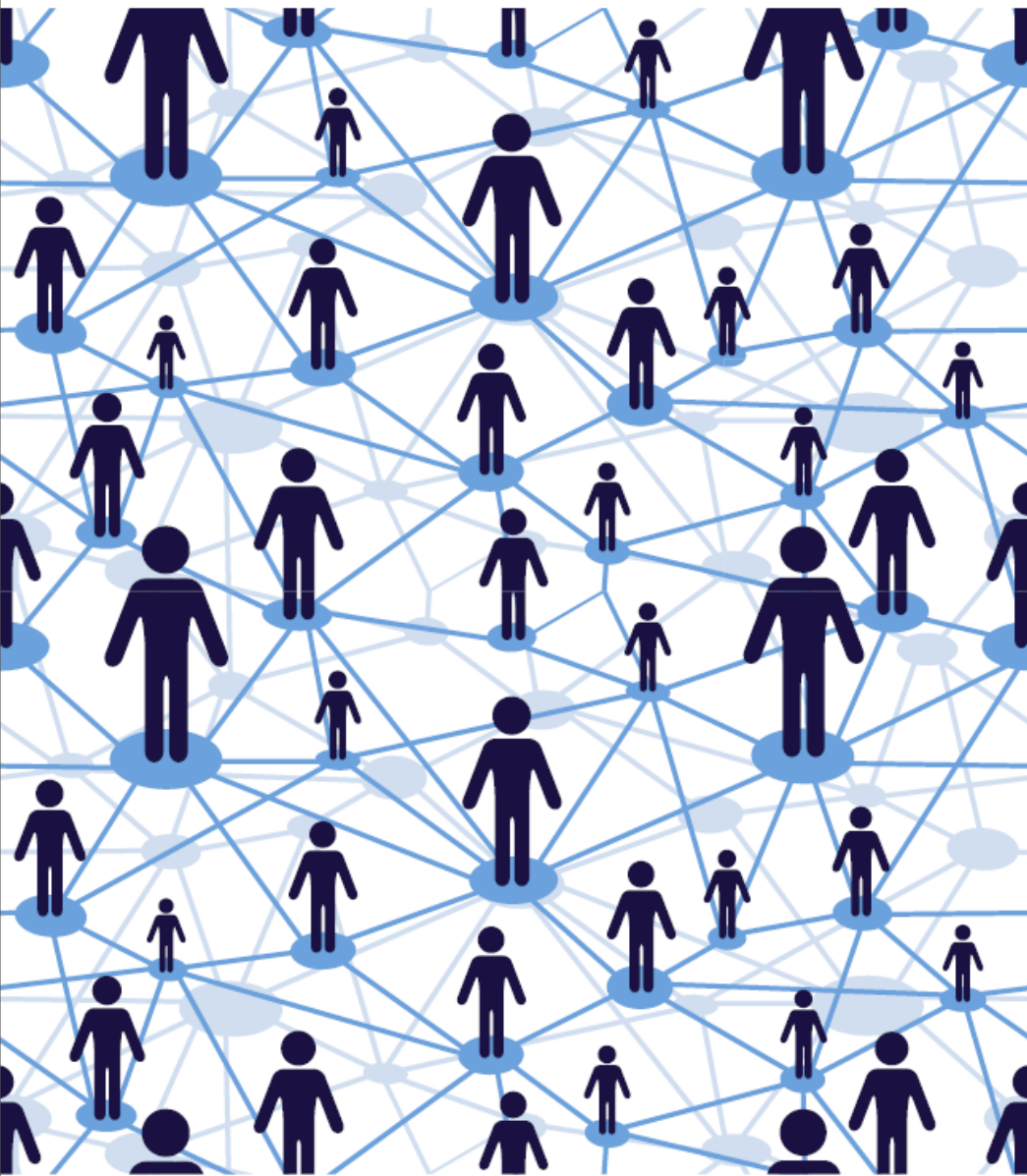
Ultimate Objective: Predictive models to navigate your health course throughout the course of your life

Adapted from Rui Chang et al. *PLoS Computational Biology*

Acknowledgements

Mount Sinai	PacBio	Cornell	CSHL
Ali Bashir Bobby Sebra Joel Dudley Andrew Kasarskis Milind Mahajan Gintaras Deikus Jun Zhu Bin Zhang Michael Linderman Gaurav Pandey Bojan Losic Omar Jabado Glenn Farrell	**Jason Chin** Yan Gao Greg Khitrov Frank Boellmann Ellen Paxinos David Rank Paul Peluso Edwin Hauw	Chris Mason Roger Altman Russell Durrett	Richard McCombie Eric Antoniou Patricia Mocombe
New York Genome Center	**Sage Bionetworks**		
Bob Darnell	Stephen Friend Chris Gaiteri		

Stay connected with us!
@multiscalebio